DEIM Forum 2023 1a-5-3

サッカーにおけるフィールドの位置推定モデルの提案

熊倉多香音† 清 雄一† 田原 康之† 大須賀昭彦†

† 電気通信大学 I 類メディア情報学プログラム 〒 182–8585 東京都調布市調布ヶ丘 1-5-1
E-mail: †kumakura.takane@ohsuga.lab.uec.ac.jp, ††sei@is.uec.ac.jp, †††{tahara,ohsuga}@uec.ac.jp

あらまし スポーツ分析においてフィールドの位置推定は多く研究されている分野である. ただし, 他のスポーツに比べてサッカーにおけるフィールドの位置推定の精度が低く, スタジアムによって異なる大きさのフィールドには対応しておらず, 様々な画角から撮られた画像に対しての位置推定の研究はあまり見られない. そこで本研究では, E^2FGVI を用いてフィールド上の人を取り除きフィールドの位置推定の精度を上げるとともに, warp error を導入することで採用するフィールドテンプレートを決定するモデルを提案した. そして, 学習において E^2FGVI またはフィールドテンプレート最適化を行った・行わなかった場合の計 4 ケースについて比較したところ, E^2FGVI 及びフィールドテンプレートの最適化を行った場合が最も精度が高くなった.

キーワード 画像, 深層学習, サッカー, スポーツ, 射影変換, フィールドの位置推定

1 はじめに

近年, 選手の動きや状況を把握して分析する, スポーツ分析の分野が活発に研究されている. スポーツ分析は, チームでの戦略・プレイヤーのパフォーマンス判断に用いるだけでなく, プレイヤーのスカウトや試合での判定に用いられる. その中でも画像中のフィールドの位置推定は活発に取り組まれている課題である. フィールドの位置推定は, サッカーやアイスホッケー, バレーボール, アメリカンフットボール, バスケットボール, テニスといった様々なスポーツに適用されている. 選手の位置やボールの位置は, GPS を選手やボールに取り付けることによって細かい動きのデータが得られているため, 画像におけるフィールドの位置推定が実現することによって, 取得した選手の位置やボールの位置を, 人手を必要とすることなく図 1 のように自動的にビデオフレームに描くことが可能になり, 選手や戦術・テクニックへの更なる理解を深めることができる.

図 1 フィールドの位置推定によって実現できる例 [2]

しかし, サッカーにおける画像中のフィールドの位置推定はいまだ技術的に解決できていない課題がある. 具体的には, 他のスポーツに比べてフィールドの位置推定の精度が劣っていること, 様々な大きさのフィールドには対応していないことが挙げられる. 特に, フィールドの大きさについては, 国際サッカー評議会 (IFAB) によりフィールドの大きさは 105 × 68m が推奨 [12] されており, ワールドカップやオリンピック等の国際試合はこの大きさのスタジアムで行われている. そして, 日本サッカー協会でも同様に, 国内での国際試合および国民体育大会等の大会でのフィールドでの大きさは 105 × 68m と定められている [27]. しかし, 海外には 105 × 68m でないスタジアムも存在する. IFAB でもフィールドの大きさは 90-120 × 45-90m の範囲であればよいと定められている [12].

そこで本研究では, フィールド上の人を取り除くことでフィールドの位置推定の精度を上げるとともに, 様々な画角やフィールドの大きさに対応できることを目的としたフィールドの位置推定を行う手法を提案する. 本研究によって, フィールドの位置推定を, 撮る方向やフィールドの大きさにとらわれることなく実現できるようになるため, 更に広くこの技術が使われることが期待できる.

本研究で新しく取り入れる点は, フィールドの位置推定を行う前にフィールド上の人を取り除くこと, フィールドの位置推定においてフィールドテンプレートの選択を行うことである. 既存研究において他のスポーツに比べてフィールドの位置推定の精度が劣っているのはプレイヤーがフィールドラインに重なっていることが原因であると考え, 本研究ではフィールドの位置推定を行う前処理として, E^2FGVI [17] を導入する, また, フィールドの位置推定において warp error を導入し, これをフィールドの位置推定の精度を高めることに用いるだけでなく, warp error が最も小さいものをフィールドテンプレートとして採用することでフィールドテンプレートの大きさを決定する, フィールドテンプレートの最適化を行う.

そして, E^2FGVI 及びフィールドテンプレート最適化を行った場合, E^2FGVI を行わずフィールドテンプレート最適化のみを行った場合, フィールドテンプレート最適化を行わず E^2FGVI を行った場合, E^2FGVI 及びフィールドテンプレート最適化を行わなかった場合の 4 つのケースについて比較した結果, E^2FGVI 及びフィールドテンプレート最適化を行った

ケースが最も平均 MAE の値が小さくなった.

本論文の構成は以下のとおりである. 2 章では関連研究, 3 章では提案手法の詳細な説明, 4 章では提案手法を用いた実験およびその結果, 最後に 5 章で考察, 6 章で本論文の結論をまとめ, 今後の展望を示す.

2 関 連 研 究

2.1 ホモグラフィ推定とカメラキャリブレーション

フィールドの位置推定を行うには主にカメラキャリブレーションとホモグラフィ推定の 2 つのアプローチが存在する.

カメラキャリブレーションとは, レンズの歪みパラメータ, イメージまたはビデオカメラのイメージセンサーのレンズ焦点距離や光学中心などの内部パラメータ, ある座標系における 3 次元点の座標を別の座標系での座標に変換するための回転と並進を表す外部パラメータを推定し, 画像を補正する処理のことである [14] [19]. 3 次元における座標とそれに対応する 2 次元における座標を用いてこれらのパラメータを推定することで. 3 次元世界での画像を 2 次元平面へとマッピングすることができる [19].

対してホモグラフィ変換とは, 射影変換により平面を異なる平面へ射影することを指す. ホモグラフィ変換は下式 (1) のようにホモグラフィ行列と呼ばれる 3 × 3 の行列を用いて変換前の座標 (x_1, y_1) から変換後の座標 (x_2, y_2) を求める.

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (1)$$

2 つの画像間に対応する 4 つの点の座標からホモグラフィ行列 H を推定することで片方の画像上の任意の点を, 対応するもう一つの画像上の点へと変換することができる [18].

2.2 フィールドの位置推定

画像中のフィールドの位置推定は様々なアプローチで研究されてきた. Watanabe ら [26] は, 画素の指定によりフィールド上の幾何学模様 (円, 直線) を抽出し. ワイヤーフレーム作成モデルと模様の一致度からカメラキャリブレーションを行うことでフィールドの位置推定を行った.

そして, ホモグラフィ推定によりフィールドの位置推定を行う手法では近年, ディープラーニング的手法 [11] [13] [21] がフィールドの位置推定の精度を高めてきた.

Homayounfar ら [11] はディープラーニングを用いて放送画像のセマンティックセグメンテーションを行い, それを用いて幾何学的な事前知識を持った Markov Random Field 上でフィールドとカメラのポーズのパラメーターを推定することで, フィールドの位置推定を行っている. 他にも, Jiang ら [11] は 2 つの DNN を用い, ホモグラフィ行列を求めるネットワークから求めた値をもとにテンプレートを変形させ, これらのペア画像を 2 つ目の DNN に入力し, 得られた Registration Error をもとにホモグラフィをアップデートするという最適化プロセスを繰り返すことによって精度を高めている. Shi ら [21] はクロスモダリティ間の異なる視点の 2 枚の画像をリグレッションネットワークに入れることで画像間のホモグラフィを推定し, iteration ごとに出力されたホモグラフィを適用したエッジ画像を再び入力として用いることで正解値に近づけていくとともに, score 値からアラインメントエラーを推定することによって iteration の回数をコントロールしながら最適化を行った. また, Detone ら [6] の手法を参考に ImageNet からの画像とランダムに歪ませた画像をリグレッションネットワークに入力し, その画像間のホモグラフィパラメータを推定するという事前学習をした上でフィールドの位置推定を行うことで精度が上がることを示している.

2.3 Mask Transfiner

Mask Transfiner とは, インスタンスセグメンテーションを行う transformer ベースのアルゴリズムである. 図 2 がそのアーキテクチャを表している.

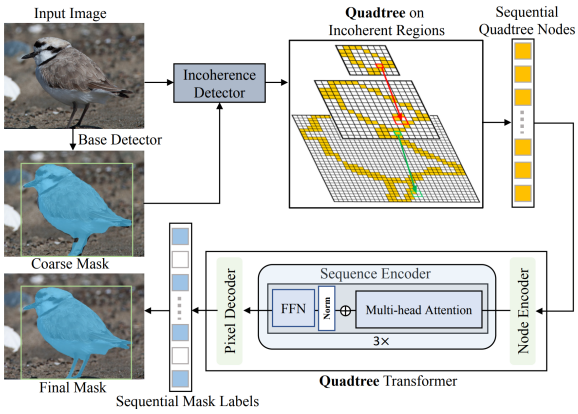

図 2 Mask Transfiner の構造 [15]

画像領域をピラミッド状にした Quadtree から得られた Sequential Quadtree nodes を Quadtree Transformer に入力として与える. Quadtree においてマスクのエラーを修正することによって, 関心領域におけるマスクの精度を向上させている [15]. 図 3 は, インスタンスセグメンテーションを行うアルゴリズムである, Mask R-CNN, SOLQ, PointRend との比較を行った例である. 図 3 から, これまでのインスタンスセグメンテーションよりマスクの精度が上がっているのが分かる.

図 3 インスタンスセグメンテーションを行うアルゴリズムの比較 [15]

2.4 E^2FGVI

映像内の物体除去を行うことができる動画修復は, 障害物除去, ビデオの復元などに実世界のアプリケーションに広く適用されているものの, 各フレームに対して直接的に画像修復を行うことで時間的に一貫性のない映像が生成されてしまうという課題を残している. E^2FGVI とは, Li ら [17] が提案した, 動画修復を End-to-End に行うフレームワークである. 入力として動画をフレームごとに分割した画像と, それに対する修復領域が示されたマスク画像を用いることで, 出力としてマスク画像で指定した領域が修復された画像を得ることができる. E^2FGVI ではフロー補完, 特徴伝搬, コンテンツ幻視の 3 つの学習モジュールで密接に連携することで高い効率性を持ちながら最先端の精度を達成している.

2.5 SoccerNet

サッカーにおけるフィールドの位置推定に用いられるデータセットとして, Soccer World Cup dataset [11], SoccerNet-v2 [5] が挙げられる. Soccer World Cup dataset [11] は, ブラジルで開催された World Cup 2014 での 20 試合 395 枚 (学習 209 枚, テスト 186 枚) の画像に対して正解ホモグラフィを求めたデータセットである. それに対して SoccerNet は 2014 年から 2017 年の EPL, ラ・リーガ, リーグ 1, ブンデスリーガ, セリエA, UEFA チャンピオンズリーグでの 500 試合のサッカー映像を提供しており, SoccerNet-v2 は SoccerNet [9] を拡張したデータセットである. SoccerNet では現在, Action Spotting, Replay Grounding, camera shot segmentation, Field localization, Camera calibration, player tracking, Player Re-Identification といった様々なタスクそれぞれに応じたデータセットを提供しており, カメラの位置や画角も様々なものが存在する.

3 アプローチ

本研究で提案するモデルは図 4 の通りである.

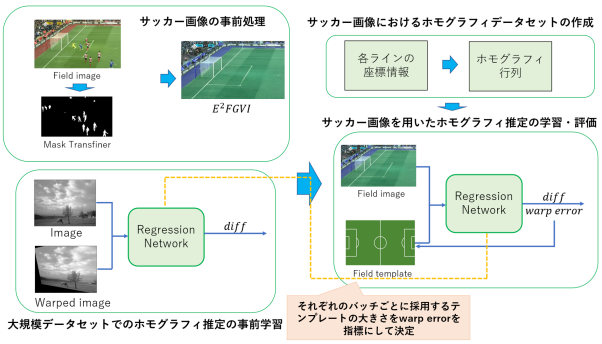

図 4 提案モデル

3.1 サッカー画像の事前処理

サッカー画像の事前処理では, E^2FGVI を用いてフィールド上のインスタンスを取り除かれた画像を得る. 本論文において E^2FGVI の入力に用いるマスク画像は Mask Transfiner [15] を用いて作成した. Mask Transfiner において, サッカー画像を入力することでオブジェクトとして認識されるのはボール及び人のみであるため, 認識されたオブジェクトからマスク画像を作成し, E^2FGVI を適用することでボール及び人を画像中から除去することができる.

3.2 サッカー画像におけるホモグラフィデータセットの作成

本研究では, 図 5 のように各フィールドイメージに対するサッカーピッチの要素の端部の座標のみから 2 次元におけるフィールドテンプレートのどの位置に値するのかを識別させ, それらの対応点からホモグラフィ行列を求める.

図 5 フィールドの位置推定を行う様子

図 6 サッカーピッチ画像の例

第 4.1 節で述べるデータセットに含まれているサッカーピッチ画像は図 6 のように, フィールドの斜め上から撮影されたものだけでなく, ゴール内から撮影された画像やプレイヤーの真横から撮影された画像, ゴールに向かって撮影された画像など, 360 度様々な画角から撮影されたものとなっている. 例えば, 図 6 の左から 3 番目の画像におけるピッチラインの座標として格納されているのは”Goal left post right”, ”Goal left cross bar”, ”Side line left”, ”Big rect. left bottom”, ”Small rect. left bottom”の端点である. これらのピッチラインは全てフィールドを横から見たときに左側に存在するピッチラインであるが, 画像のみから判断するとフィールドを横から見たときの右側を映している. つまり, それぞれの端点がフィールドテンプレートのフィールドラインにおける端点のどちらに属しているのかを x, y 座標の大小のみで定めるのは誤った変換である.

そこで本研究では図 7 のように, フィールドテンプレートを横から見たとき平行な線, 垂直な線, ペナルティーアークの 3 つに場合分けする.

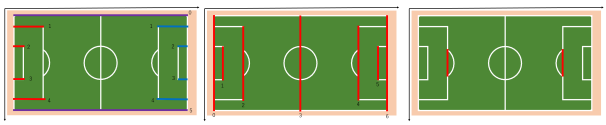

図 7 フィールド端部の座標からホモグラフィ行列を求める場合分け

まずフィールドテンプレートを横から見たとき平行な線について, フィールドテンプレートを横から見たときの左右に 2 グループに分割した. タッチラインについては両方のグループに属している. そしてそれぞれのグループにおいて, 与えられた

直線 l_1 及び l_1 と比べる直線 l_2 を定め, それぞれの直線の傾きを計算する.

直線 l_1 の傾き m_1 が $-1 < m_1 < 1$ であるとき l_1, l_2 の端点のうち x 座標が小さい方の x を基準として, 基準 x における y 座標を求める. そしてその y 座標の値が小さい方の直線がフィールドテンプレートにおける上のラインであるとき正面から撮っていると判断し, それぞれの直線の x 座標が小さい方をフィールドテンプレートのラインにおける”left”, x 座標が大きい方を”right”とする. そうでないとき裏側から撮っていると判断し, それぞれの直線の x 座標が大きい方を”left”, x 座標が小さい方を”right”とする.

同様に, $m_1 <= -1$ または $1 <= m_1$ であるとき l_1, l_2 の端点のうち y 座標が小さい方の y 座標を基準として, 基準 y における x 座標を求める. そしてその x 座標の値が小さい方の直線がフィールドテンプレートにおける上のラインであるとき left 側から撮っていると判断し, それぞれの直線の y 座標が大きい方を”left”, y 座標が小さい方を”right”とする. そうでないとき right 側から撮っていると判断し, y 座標が大きい方を”right”, y 座標が小さい方を”left”とする.

またフィールドテンプレートを横から見たとき垂直な線についても同様に, 与えられた直線 l_1, 比べる直線 l_2 を定め, それぞれの直線の傾きを計算する.

直線 l_1 の傾き m_1 が $-1 < m_1 < 1$ であるとき l_1, l_2 の端点のうち x 座標が小さい方の x 座標を基準として, 基準 x における y 座標を求める. そしてその y 座標の値が小さい方の直線がフィールドテンプレートにおける右のラインであるとき left 側から撮っていると判断し, x 座標が小さい方を”top”, x 座標が大きい方を”bottom”とする. そうでないとき right 側から撮っていると判断し, x 座標が小さい方を”bottom”, x 座標が大きい方を”top”とする.

直線 l_1 の傾き m_1 が $m_1 <= -1$ または $1 <= m_1$ であるとき l_1, l_2 の端点のうち y 座標が大きい y 座標を基準として, 基準 y における x 座標を求める. そしてその x 座標の値が小さい方の直線がフィールドテンプレートにおける左のラインであるとき正面から撮っていると判断し, y 座標が大きい方を”bottom”, y 座標が小さい方を”top”とする. そうでないとき裏側から撮っていると判断し, y 座標が大きい方が”top”, 小さい方が”bottom”とする.

そしてペナルティーアークにおける端点においては, フィールドテンプレートを横から見たとき平行なライン, 垂直なラインの位置関係を求める際に求めたカメラの方向と位置をもとに位置関係を定める. 正面から撮っている場合, y 座標が大きい方が”bottom”, y 座標が小さい方が”top”であり, 裏側から撮っている場合, y 座標が大きい方が”top”, y 座標が小さい方が”bottom”である. また left 側から撮っている場合, x 座標が大きい方が”bottom”, x 座標が小さい方が”top”であり, right 側から撮っている場合 x 座標が大きい方が”top”, x 座標が小さい方が”bottom”である.

このとき, ホモグラフィ推定に用いる 4 点のうち 3 点以上が同一直線上に存在するとホモグラフィ行列が正しく推定できないため, 3 点以上が同一直線上に並ばないよう, 選択した点同士の傾きを計算して閾値以内の誤差となった直線が存在した場合, そのうち片方の点をホモグラフィ推定に用いないこととする. そしてその結果まだ同一直線上に並んでいる 3 点を採用してホモグラフィ推定を行った画像については手作業で削除する.

また, ホモグラフィ推定に用いる端部の座標が画像内に存在しない場合, データセットに含まれる端部の座標は正規化されているため, 0 未満または 1 より大きい値となっている. また 0 以上 1 以下の値であっても画像の端に近い箇所では明らかにフィールドラインの端点となっていない部分が存在するため, x 座標, y 座標ともに 0.01 以上 0.99 以下となっている点を採用する. そして, 指定された座標が”Big rect. left top”の”right”と”Big rect. left main”の”top”のようにフィールドテンプレート上で重なる点であった場合, これらの座標の中点を採用する. そしてカメラの方向や位置 (正面, 裏側, left 側, right 側) が計算している時点で異なる結果となった場合はその画像は採用しないこととする. センターサークルについては端点ではなく格納されている座標をつなげることで円になる座標が格納されており, 今回は端点に注目するため採用しないこととする.

これを全てのフィールドラインの端点に対して行い, ホモグラフィ変換に用いる 4 点が決まった時点でその 4 点を用いてホモグラフィ行列を求め, フィールドテンプレートにおけるタッチラインの端点 4 点に対して求めたホモグラフィ行列を適用することで変換後のタッチラインの端点の座標, つまりフィールド画像におけるタッチラインの端点の座標を求める.

3.3 事前学習におけるデータの前処理

事前学習に用いる大規模なデータセットには Detone ら [6] と同様のデータの前処理を行う. 図 8 のように各画像に対してランダムな位置に 128 × 128 のパッチ画像 A を作成し, そのパッチ A の 4 隅を [-32, 32] の範囲内で歪ませる. そしてこれらの対応点のホモグラフィ行列 H を求め, その逆行列を画像に適用させることで得られた画像をパッチ B とする. このパッチ A, B と各対応点の差分を 1 つのペアとしてデータセットを作成する.

図 8　事前学習におけるデータの前処理

3.4 ネットワークの構造

Shi ら [21] の研究では, 事前学習により位置推定の精度が上がっていたため, 本研究においてもその手法を採用する. 具体的なネットワークとして Detone ら [6] のネットワークを用いて実験を行う. そしてサッカー画像の学習において, このネットワークの出力に対し warp error を導入することで採用するフィールドテンプレートの大きさを決定する. Warp error は前回の warp error を超えた場合, その前の時点でのフィールドテンプレートの大きさを採用するとともに, その前の時点での loss

を用いてネットワークのパラメータを更新する．ここで，warp error は次式 (2) で比較して求める．MAE とは正解値と予測値から求まる平均絶対値誤差のことであり，重み pdf_{pre} として追加しているのが前回モデルに入力したテンプレートの確率密度である．バッチごとに MAE を初期値 inf，pdf_{pre} を初期値-1 とすることで最低でも 1 回は確率密度が最も高いテンプレートを選ぶように設定している．

$$warp\ error = MAE \times pdf_{pre} \tag{2}$$

また，ネットワークで予測するのは指定した対応点 4 点の座標の差分である．事前学習において予測するのはバッチ A，B の各対応点である端点の差分であり，サッカー画像の学習において予測するのはフィールドテンプレートとサッカー画像間の対応点であるタッチラインの端点 4 点の差分である．

3.5　サッカー画像における学習

様々な大きさのフィールドに対応させるために，それぞれのサッカー画像に対してフィールドテンプレートを全て試してその中の最も warp error が小さいものを採用することで，実行時間が長くなり実際のフィールドサイズとは誤ったサイズで学習することを防ぐため，本研究ではバッチサイズごとにフィールドテンプレートを決定するとともに，サッカー画像に対するフィールドテンプレートの優先度はタッチラインとゴールラインを軸とした 2 次元正規分布の確率密度が高い順とする．2 次元正規分布を用いて確率値が高い順にフィールドテンプレートを採用することで，最もフィールドテンプレートの大きさとして可能性のあるものから順に試すことができる．

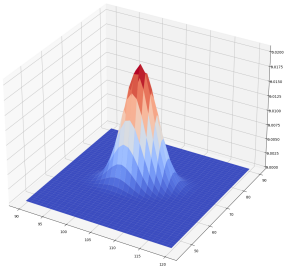

図 9　タッチラインとゴールラインを軸とした 2 次元正規分布

4　実　　験

4.1　使用したデータセット

本研究では，事前学習には COCO dataset を，サッカー画像の学習には SoccerNet-v2[4] における Field localization のデータセットを使用する．Field localization タスクにおけるデータセットには 20028 枚の画像と，各画像中におけるサッカーピッチ要素の 2 つ以上の端部の正規化された座標が格納されている．サッカーピッチ要素とは，フィールドラインまたはゴールポストのことであり，端部はライン/サークルの弧の端または画像末端との交点のいずれかのことである．

今回 World Cup dataset ではなく SoccerNet-v2 を採用したのは，様々なフィールドの大きさに対応させるためには異なる大きさのフィールドでプレイしている画像が必要であったが，World Cup では規定のフィールドサイズがあり，様々な大きさのフィールドでの学習には適していないためである．

4.2　フィールドテンプレートの作成

IFAB においてフィールドの大きさは 90-120 × 45-90m の範囲であればよいと定められている [12] ため，[13] のコードを参考に，90-120 × 45-90m のフィールドテンプレート画像計 1426 枚を作成した．図 10 は 105 × 68m のフィールドテンプレートを作成した様子である．

また，サッカー画像におけるホモグラフィの作成のため，それぞれフィールドテンプレートにおける 4 隅の座標を取得した．

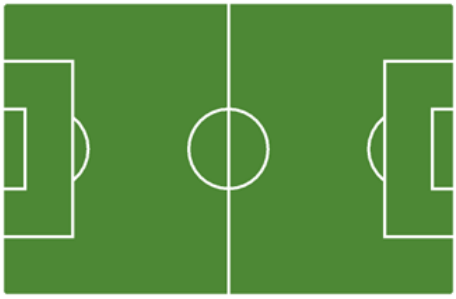

図 10　105 × 68 のフィールドテンプレート

4.3　ホモグラフィ行列データセットの作成

各フィールドラインの端部の座標のみからフィールドテンプレートのどの位置に値するのかを識別させ，それらの対応点から正解ホモグラフィ行列を求めた．そしてその正解ホモグラフィ行列の逆行列をフィールドテンプレート全体に適用することでフィールドテンプレートを変換させた．

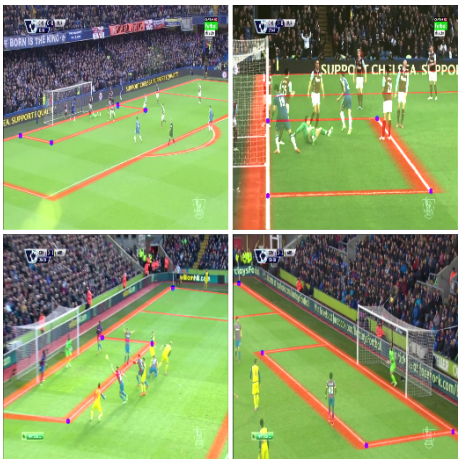

図 11　正解ホモグラフィ行列が正しく求められている例

図 11, 12 は，フィールドイメージの上に，それに対応する正解ホモグラフィ行列を適用したフィールドテンプレートを重ねて表示したものである．赤線が正解ホモグラフィ行列を適用したフィールドテンプレートを，青点がホモグラフィ変換に用い

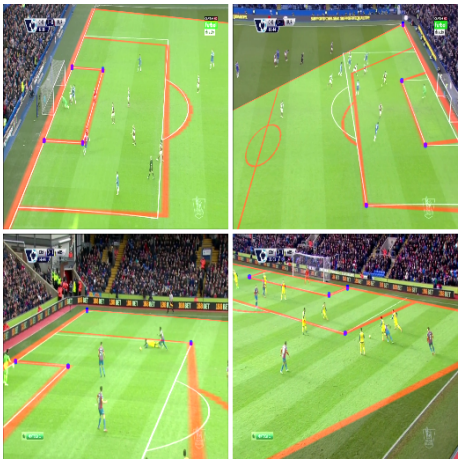

図 12　正解ホモグラフィ行列が正しく求められていない例

た点を示している. 図 11 のように正しく変換されているものもある一方, 図 12 のように, 各フィールドラインの端点はしっかりと認識されており, フィールドテンプレートにおけるフィールドラインの端点も画像内のフィールドラインの端点と一致しているものの, テンプレート全体を見るとテンプレートが大きくずれているものも存在した.

また, SoccerNet データセットからホモグラフィ行列データセットとして作成した結果, トレーニング画像 8015 枚, バリデーション画像 1482 枚, テスト画像 1443 枚となり, 各試合の対戦カード, 日付から調べた, その画像において使用されていたフィールドの大きさの内訳は表 1 のようになった [1] [3] [4] [7] [8] [10] [16] [20] [22] [23] [24] [25].

表 1　タッチライン (w) ×ゴールライン (h) の枚数

w × h	学習画像	検証画像	テスト画像
105 × 68	6235	1228	1241
103 × 68	427	35	66
101 × 68	381	43	34
105 × 66	326	80	0
110 × 68	262	14	37
105 × 70	117	0	0
100 × 68	94	0	0
105 × 69	74	0	0
104 × 74	52	0	0
110 × 70	29	0	0
100 × 67	18	29	0
105 × 71	0	32	0
108 × 70	0	21	0
105 × 65	0	0	65
計 (枚数)	8015	1482	1443

4.4　E^2FGVI の適用結果

フィールドイメージに対して E^2FGVI を行った. 図 13 は, ある画像に対して Mask Transfiner を用いてマスク画像を生成し, それらを入力として E^2FGVI を適用した様子である. Mask Transfiner の結果を見ると, 人の足の間まではっきりと認識されており, E^2FGVI の結果を見るとオリジナル画像では人によって隠れていたフィールドラインが修復されていることが分かる.

図 13　E^2FGVI の適用結果

4.5　事 前 学 習

COCO dataset を 3.1 で述べたようなアルゴリズムで加工することでホモグラフィデータセットを作成し, ホモグラフィ推定の事前学習を行った. また Detone ら [6] のネットワークは VGGNet をベースとしたネットワークであり, 入力は 128 × 128 × 2, 2 層ごとに 2 × 2 の最大値プーリング層 (ストライド 2) を入れた 8 層の畳み込み層を持っている. 畳み込み層のフィルターは, 64, 64, 64, 64, 128, 128, 128, 128 となっている. 畳み込み層の後に 2 層の全結合層があり, 第 1 層の全結合層のユニット数は 1024 である. また, 最後の畳み込み層の後と最初の全結合層の後にドロップアウト確率 0.5 のドロップアウト層を挿入し, 活性化関数は ReLU 関数である.

ここで, トレーニング画像の枚数は 118287, バリデーション画像の枚数は 40670, テスト画像の枚数は 5000 である. 学習における最適化手法として, モーメンタム 0.9 の SGD を採用し, 学習率 0.005, epoch が 16 の倍数になるごとに学習率を 10 のファクターで減衰させるよう設定した. また, 損失関数は平均二乗誤差, バッチサイズは 64, エポック数は 48, 画像サイズは 128 × 128 とした. 事前学習における loss の推移は図 14 のようになった.

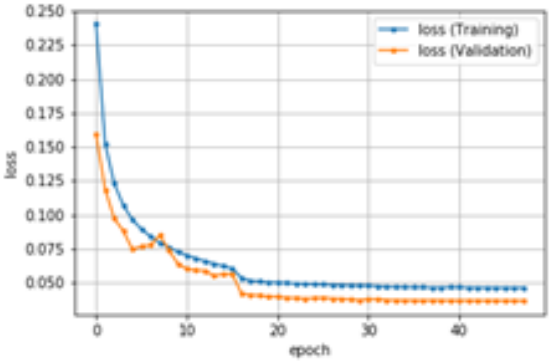

図 14　事前学習における loss の推移

4.6　サッカー画像における学習及び評価

COCO dataset で事前学習したネットワークを用いてサッカー画像の学習を行った.

学習における最適化手法として, 学習率 0.001 の Adam を採用し, epoch が 17 の倍数になるごとに学習率を 10 のファクターで減衰させるよう設定した. また, 損失関数は SmoothL1Loss, バッチサイズは 64, エポック数は 51, 画像サイズは 128 × 128

とした．また，活性化関数を事前学習時の ReLU 関数から Exponential Linear Unit(ELU) 関数へと変更した．そして 2 次元正規分布の平均値は $(105, 68)$，共分散行列は $((9, 3), (4, 10))$ とした．学習における loss の推移は図 15 のようになった．図 15 左が学習時の loss を，図 15 右が検証時の loss を表している．かなり早い段階に loss が減り，その後はあまり変化が見られなかった．

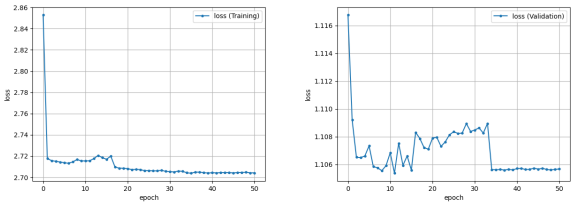

図 15　サッカー画像の学習における loss の推移

また，E^2FGVI を行った画像でフィールドテンプレートを選択する学習を行った場合，E^2FGVI を行わなかった画像でフィールドテンプレートを選択する学習を行った場合，E^2FGVI を行った画像でフィールドテンプレートを選択せず常に $(105, 68)$ のフィールドテンプレートで学習を行った場合，E^2FGVI を行わなかった画像でフィールドテンプレートを選択せず常に $(105, 68)$ でのファイルテンプレートで学習を行った場合の 4 つのケースについてテスト画像における平均 MAE を求めた結果が表 2 である．最も平均 MAE が小さかったのは E^2FGVI を行った上でフィールドテンプレートの最適化を行ったケースであった．

表 2　それぞれの場合での平均 MAE

ケース	平均 MAE
E^2FGVI あり，フィールドテンプレートの最適化あり	**379.8885**
E^2FGVI なし，フィールドテンプレートの最適化あり	394.1644
E^2FGVI あり，フィールドテンプレートの最適化なし	440.2585
E^2FGVI なし，フィールドテンプレートの最適化なし	383.6093

また，画像によって大きく MAE の値が変化していた．図 16 は，E^2FGVI 及びフィールドテンプレートの最適化を行ったとき，テスト画像の中で MAE の値が最も小さい 1.9531 であった画像であり，同様に図 17 は，E^2FGVI 及びフィールドテンプレートの最適化を行ったとき，テスト画像の中で MAE の値が最も大きい 222948.2990 であった画像である．

また，表 3 より MAE の統計量をみると，標準偏差及び分散の値が大きく，中央値が平均 MAE よりかなり小さくなったことから，画像ごとの MAE の値が大きく異なっており，中央値は平均 MAE より大きく下回っていたことから，平均 MAE の値が大きいことについて，MAE の値が非常に大きい画像の影響があることが分かる．

また，学習・検証時に採用したフィールドテンプレートの大きさの推移は図 18, 19 のようになった．図 18 において縦軸の値に 45 を足した数がその時点で採用したゴールラインの値であり，同様に図 19 において縦軸の値に 90 を足した数がその時

図 16　MAE の値が最も小さい画像

図 17　MAE の値が最も大きい画像

表 3　E^2FGVI 及びフィールドテンプレートの最適化を行った場合の MAE の統計量

統計量	値
MAE の標準偏差	6095.7137
MAE の分散	37157725.06321
MAE の中央値	55.0520

点で採用したタッチラインの値である．図 18, 19 における横軸の epoch 数はバッチごとの学習を示しているため，実際に学習した epoch 数である 51 とは異なる値となっている．

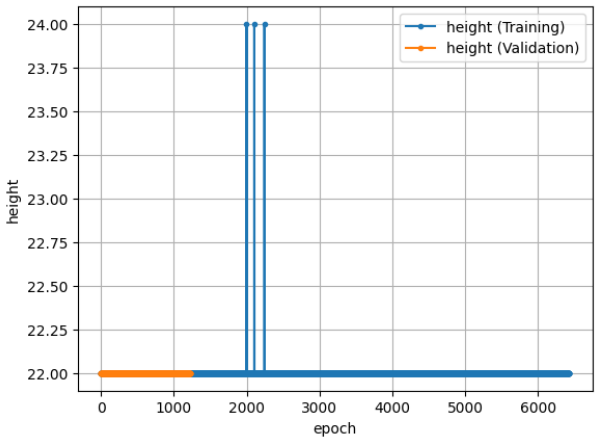

図 18　ゴールラインの推移

5　考　　察

最も平均 MAE が小さかったのは E^2FGVI を行った上で

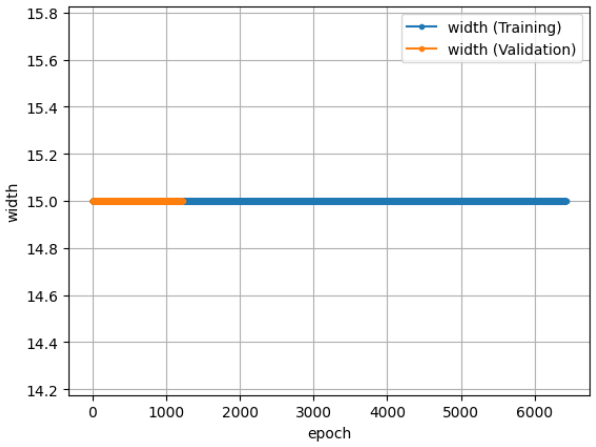

図 19　タッチラインの推移

フィールドテンプレートの最適化を行ったケースであったことから, E^2FGVI を行った上でフィールドテンプレートの最適化を行うケースが最も精度が良かったことが分かる. しかし, このケースにおいて精度は上がったものの, フィールドテンプレートの最適化を行わなかった場合 E^2FGVI を行うことで逆に精度が落ちるという結果となっていたため, E^2FGVI を行うことによって精度が非常に上がったとは言えない結果となった. これは E^2FGVI を行うことによって画像の特徴量が減り, ネットワークが学習できなかったとも考えられる.

また, 次いで精度が良かったものが E^2FGVI をせずにフィールドテンプレートの最適化もしなかったケースであった. これはテスト画像のフィールドテンプレートの大きさがほぼ 105×68 となっており, その他のフィールドの大きさがかなり少なくなっていたことが影響したと考えられる. また, 図 18, 19 でみられるように学習・検証時のタッチライン及びゴールラインの変化があまり見られなかった. 本研究での学習・検証画像のフィールドの大きさはさまざまなフィールドの大きさのものが存在したものの, 105×68 のものが非常に多かったため, タッチライン及びゴールラインの変化が時折見られたという結果は自然なものであったが, 本研究の有効性をより確かめるためには, データ数・サイズ数ともにより多くのフィールドサイズで学習・評価することが必要だと考えられる. これによりフィールドテンプレートの最適化を行った場合と行わなかった場合の差が大きくなるとともにフィールドサイズの最適化が変化する様子が観察できると期待できる.

そして, 図 16, 17 のように MAE の値が小さかった画像と大きかった画像を比べると, フィールドの大きさや画角ではなく画像全体が暗く, フィールドラインが際立っている画像の MAE が小さく, 光と影が画像内にできている画像が最も MAE が大きかったことで, 光と影の境界線がフィールドライン特徴量として捉えられたのではないかと考えられる. そこで, 光と影のコントラストを抑えることで MAE の値を小さくできると期待できる.

また本研究での平均 MAE が非常に大きくなった理由としては複数考えられる.

まず, 事前学習では画像の 4 隅のずれが最大 32 であったのに対し, サッカー画像における 4 隅は画像に映らないケースが多かったため, 画像のサイズである 128 以上となることも多かったためだと考えられる. この対策として, 32×32 より大きいずれを作ったデータセットで事前学習することが挙げられる.

次に, 表 3 で示したように, 画像ごとの MAE の値が大きく異なる結果となったことが影響していると考えられる. MAE の値が非常に大きくなった画像の影響で平均 MAE が大きくなっているため, データ数を増やして学習・及び評価を行うことで MAE の値が非常に大きいものの影響が出にくくなると考えられる.

また, SoccerNet の Field localization タスクで提供されていたピッチラインの座標からホモグラフィ行列を求めたが, 図 12 で示したようにホモグラフィ行列を適用した結果がずれてしまい, 目視で見ても真の値とは言えないものが多く存在した. 本研究においてはネットワークで予測したのはサッカー画像におけるタッチラインの端点 4 点であるため, フィールドが少しでもずれることによって画像外の端点の座標情報に大きく影響を及ぼしたと考えられる.

そしてネットワークの構造を変更していく必要があったと考えられる. 本論文ではサッカー画像における学習率や最適化手法を 4.6 節のように定めたが, 学習率や最適化手法, 活性化関数を変化させて本研究での実験を行ったところ, それぞれの画像ごとの MAE の差が非常に大きくなっており, loss が減りづらいという問題点が存在する. そのため, 学習率や最適化手法, 活性化関数だけでなくネットワークの構造をよりディープにする必要があると考えられる.

6　まとめ及び今後の展望

本論文では, E^2FGVI を用いてフィールド上の人を取り除くことで精度を上げるとともに, SoccerNet データセットから得られる座標情報からホモグラフィ行列を定義し, フィールドの大きさを指定することで, 様々な画角やフィールドの大きさに対応できることを目的としたフィールドの位置推定を行うシステムを提案した. サッカー画像における学習は考察で述べたような改善点を解決することによってより精度の高いものができるとともに, より本研究の有効性を確かめることができると考えられる.

今後の展望として, 事前学習におけるデータセットの工夫の改善, データセットの多様化, ホモグラフィ行列の定義の洗練化, ネットワークの構造の変更, 光と影のコントラストを抑えることが考えられる. 特にデータセットの多様化については, データセットの数を増やすとともにフィールドサイズの種類をより増やすことが求められるが, 本研究においてはホモグラフィ行列定義を行った結果, サッカーデータセットの数が減ってしまったため, pix2pix を用いて様々なフィールドの大きさのサッカー画像を生成することでデータセットの拡張が可能となり, より高い精度の結果が得られるとともに, 様々なフィールドの大きさに対応しやすくなり, 本研究の手法の有効性をより確かめら

れるのではないかとも考える.

7 謝　　辞

本研究は JSPS 科研費 JP21H03496, JP22K12157 の助成を受けたものです．また本研究は，電気通信大学人工知能先端研究センター (AIX) の計算機を利用して実施したものです.

文　　献

[1] BeSoccer, (Accessed on 1/2023). https://www.besoccer.com/.

[2] Bundesliga. Dfl and amazon web services to provide new real-time match analysis, 2019.

[3] Valencia CF, (Accessed on 1/2023). https://www.valenciacf.com/.

[4] World Stadium Database. https://www.worldstadiumdatabase.com/, (Accessed on 1/2023).

[5] Adrien Deliège, Anthony Cioppa, Silvio Giancola, Meisam J. Seikavandi, Jacob V. Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B. Moeslund, and Marc Van Droogenbroeck. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4503–4514, 2021.

[6] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation, 2016.

[7] Football Fandom, (Accessed on 1/2023). https://football.fandom.com/wiki/Football_Wiki.

[8] FIFPlay, (Accessed on 1/2023). https://www.fifplay.com/.

[9] Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. Soccernet: A scalable dataset for action spotting in soccer videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1792–179210, 2018.

[10] GOALZZ, (Accessed on 1/2023). https://www.goalzz.com/.

[11] Namdar Homayounfar, Sanja Fidler, and Raquel Urtasun. Sports field localization via deep structured models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[12] IFAB. Law1 the field of play. https://www.theifab.com/laws/latest/the-field-of-play/#field-surface, (Accessed on 29/12/2022).

[13] Wei Jiang, Juan Camilo Gamboa Higuera, Baptiste Angles, Weiwei Sun, Mehrsan Javan, and Kwang Moo Yi. Optimizing through learned errors for accurate sports field registration. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.

[14] Alexander Mordvintsev Abid K. カメラキャリブレーション, (Accessed on 10/011/2023). http://labs.eecs.tottori-u.ac.jp/sd/Member/oyamada/OpenCV/html/py_tutorials/py_calib3d/py_calibration/py_calibration.html.

[15] Lei Ke, Martin Danelljan, Xia Li, Yu-Wing Tai, Chi-Keung Tang, and Fisher Yu. Mask transfiner for high-quality instance segmentation, 2021.

[16] LaLiga, (Accessed on 1/2023). https://www.laliga.com/en-GB.

[17] Zhen Li, Cheng-Ze Lu, Jianhua Qin, Chun-Le Guo, and Ming-Ming Cheng. Towards an end-to-end framework for flow-guided video inpainting, 2022.

[18] Satya Mallick. Homography examples using opencv (python / c ++). https://learnopencv.com/homography-examples-using-opencv-python-c/, (Accessed on 10/01/2023).

[19] MathWorks. What is camera calibration?, (Accessed on 10/01/2023). https://jp.mathworks.com/help/vision/ug/camera-calibration.html?lang=en.

[20] Playmakerstats, (Accessed on 1/2023). https://www.playmakerstats.com/home.php.

[21] Feng Shi, Paul Marchwica, Juan Camilo Gamboa Higuera, Michael Jamieson, Mehrsan Javan, and Parthipan Siva. Self-supervised shape alignment for sports field registration. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3768–3777, 2022.

[22] Football Stadiums. https://www.football-stadiums.co.uk/#:~:text=Football%20Stadiums%201%20Champions%20League%20Stadiums%20Santiago%20Bernab%C3%A9u,Stadiums%20...%207%20Stadiums%20Around%20The%20World%20, (Accessed on 1/2023).

[23] TransferMarkt. https://www.transfermarkt.jp/, (Accessed on 1/2023).

[24] Football Tripper, (Accessed on 1/2023). https://footballtripper.com/.

[25] virtual globetrotting, (Accessed on 1/2023). https://virtualglobetrotting.com/.

[26] Tomoki Watanabe, Miki Haseyama, and Hideo Kitajima. A soccer field tracking method with wire frame model from tv images. *2004 International Conference on Image Processing, 2004. ICIP '04.*, Vol. 3, pp. 1633–1636 Vol. 3, 2004.

[27] 公益財団法人日本サッカー協会. サッカー競技規則 2021/22, (Accessed on 29/12/2022). https://www.jfa.jp/laws/soccer/2021_22/.