

大規模空間における動的シーン状況把握のための 多視点映像切替手法

岡田庸佑^{†1} 糸戸英彦^{†2} 亀田能成^{†2} 北原格^{†2}

概要: 本研究では、動的シーンの状況把握に適した多視点映像切替を実現するために、選手、ボール、ゴールの位置情報に基づいて、シーンの状況把握に適した多視点映像の自動切替手法を提案する。計算能力の向上やカメラの高画質化により、サッカースタジアムなどの大規模空間を多視点から撮影した映像が利用可能になりつつある。被写体を多面的に観察できる多視点映像を活用することで、新しい発見や状況のさらなる理解が可能となることが期待される。多視点映像提示手法の一つである Bullet-Time 映像は、画質を保ちながら多視点映像の閲覧が可能であるため、被写体の詳細な観察に適した手法である。従来の Bullet-Time 映像では、静止シーンの観察を対象としているが、サッカーの試合のような動的なイベントを多面的に観察する場合、時間経過とともに変化する被写体の状況に応じた Bullet-Time 映像生成処理が必要とされる。本稿では、大規模空間としてサッカースタジアムを、動的シーンとしてサッカーの試合を、多視点映像提示手法として Bullet-Time 映像に着目し、選手、ボール、ゴールの位置情報に基づいて、Bullet-Time 映像を生成するためのパラメータ（注視点、観察視点、ズーム値）の推定法について述べる。

キーワード: 多視点映像, Bullet-Time 映像, 映像切替, カメラワーク, サッカー映像

1. はじめに

計算機処理能力の発展や高画質な映像の撮影が可能なカメラの普及により、サッカースタジアムのような大規模空間で撮影した多視点映像の活用が活発になりつつある[1-3]。大規模空間の映像を観察する場合には、被写体を多面的に観察できる多視点映像を活用することで、単一のカメラで撮影した映像では気づかない事象発見やさらなる状況把握が期待できる。その特性に基づき、複数台のカメラで撮影した映像や3次元モデルから任意視点の被写体の見え方を再現する研究が盛んに行われている[4-6]。

多視点映像提示手法の一つに Bullet-Time 映像がある。Bullet-Time 映像の長所は、撮影した画像をほぼそのままの状態では提示するため、画像変換処理による画質劣化の影響を受けにくいことであり、被写体の詳細な観察に適した手法である。サッカーの試合を多面的に観察するためには、時間経過とともに被写体群の位置が変化する動的シーンに対応した多視点映像生成が必要となる。動的シーンに対応した Bullet-Time 映像を生成するためには、撮影空間中の何処に注目するか（注視点）、どの方向から観察するか（観察視点）、どのくらいのサイズで観察したいか（ズーム値）に関する映像切替パラメータを動的に設定する必要がある。従来の Bullet-Time 映像生成法[7]では、マウスなどの入力装置を用いた手動操作によって注視点、観察視点、ズーム

値を設定するため、多面的に観察したい瞬間が訪れる度に、観察者は映像を一時停止して映像切替パラメータを設定する必要があった。その上、映像切替パラメータを動的かつ適切に設定し続けることは、多視点映像生成アルゴリズムや映像制作に関する知識が必要とされるため、一般の視聴者にとっては瞬時に適切な設定を行うことが難しい。

本研究では、大規模空間で行われる動的イベントへの適用可能な Bullet-Time 映像生成法（具体的には多視点映像の自動切替手法）を実現し、状況把握に適した多視点映像の自動生成を目指す。大規模空間としてサッカースタジアムを、動的シーンとしてサッカーの試合に着目し、サッカーの状況を表す重要な情報源である選手、ボール、ゴールの位置情報を活用することで、映像切替パラメータを推定する。

2. 関連研究

2.1 Bullet-Time 映像

Bullet-Time 映像は、多視点映像提示手法の一つであり、画質の低下を防ぎつつ、視点の移動感を再現可能という特徴を有する手法である。Akechi らは、注視点を移動可能な Bullet-Time 映像生成手法を提案した[7]。Ikeya らは、ロボットカメラで撮影した多視点映像を活用し、固定されたカメラに比べてより広い空間を観察できる Bullet-Time 映像生成に成功している[8]。これらの手法では、注目したいタ

^{†1} 筑波大学 大学院システム情報工学研究科
Graduate School of Systems and Information Engineering,
University of Tsukuba

^{†2} 筑波大学 計算科学研究センター
Center for Computational Sciences, University of Tsukuba

イメージを切り出したシーン（静止シーン）の画像を用いて Bullet-Time 映像を生成しており、シーンの状況変化に対応した映像生成の自動化は実現されていない。サッカーなどの常に選手やボールが移動する動的シーンの観察に Akechi らの手法を適用する際には、映像切替パラメータである注視点、観察視点、ズーム値を手動で設定する必要があるため、膨大な労力と作業時間を要する。本研究では、動的シーンの全てのフレームにおいて、試合状況の観察に適した映像切替パラメータを自動推定し、動的シーンに適応した Bullet-Time 映像の自動生成を目指す。

2.2 多視点映像の適切な視点設定

多視点映像の観察では、複数台の多視点カメラから観察に適したカメラを1台選択する必要がある。Chen らは、複数の映像からライブ放送で使用する映像を自動でランク付けする研究を行った[9]。ホッケーの試合を対象に、フィールド場の選手位置情報と各カメラの位置情報を活用し、映像監督者が選んだ映像を推定している。Jiang らは、複数の被写体追跡を行い、最適な視点選択を再帰的決定問題として定式化した[10]。この研究では、動き回る3人の人物を撮影した三つの映像から最適な視点を選ぶ実験を行っており、Bullet-Time 映像のように多数のカメラで撮影した映像切替設定への適用は考えられていない。本研究では、サッカースタジアムに Bullet-Time 映像の生成可能な多数台のカメラによる撮影を行うことに加え、被写体（サッカー選手）が22人以上になる大規模空間特有の課題について検討し、そのような状況に対応可能な視点選択手法を提案する。具体的には、サッカーの試合状況を理解するために重要な情報源である選手、ボール、ゴールの位置情報を活用し、映像切替パラメータを推定する。

2.3 選手やボールの位置推定

画像中の被写体の位置情報を推定する研究が盛んに行われている。Iwase らは、複数台のカメラで観測されるサッカー選手領域から、その3次元位置を推定している[11]。糟谷らは、影情報を活用することにより、2台のカメラで撮影した映像からサッカーフィールド上の選手位置を安定して推定する手法を提案している[12]。

本研究では、Bullet-Time 映像を生成することを目的としており、観察する被写体を取り囲むように多数のカメラが配置されているため、関連研究と比べ桁多い台数のカメラが利用可能である。それらのリッチな映像情報を活用することにより、選手同士でのオクルージョンの影響を軽減し、より頑健な選手位置推定の実現を目指す。

サッカーボールの位置推定を行う研究も盛んに行われている。Ren らは、複数台のカメラを使用し、フィールド上でのボールの3次元位置を推定している[13]。石井らは、

固定された二つのカメラで撮影した映像からサッカーボールを検出し、ボールの3次元位置推定を行なっている[14]。選手などによってボールが隠れてしまう場合、ボール位置の補間処理を行う際、映像中でボールが十分に観測されていない部分で途切れた軌道になり、適切な予測に失敗している。

Bullet-Time 映像を生成する際には、同じエリアを撮影した多数の映像を活用するため、複数の視点でボールを観測することが可能である。その結果、一般的な被写体検出法でも、より多くのフレームでボールの検出が可能である。本稿では、その特性を活かしたロバストな選手とボールの追跡処理について述べる。

2.4 スポーツ映像を用いた状況把握支援

スポーツ映像を用いた状況把握に関する取り組みが盛んに行われており、テレビ放送で使用するコンテンツの提供だけでなく、戦術やパフォーマンスの分析に映像情報が活用されている[15]。樋口らは、サッカーの試合ごとのパフォーマンスとその変遷を可視化するために、実際の試合をスタンド内上部よりパンニング撮影した映像を活用し、攻撃場面の抽出等を行っている[16]。実験中には、試合映像を活用したミーティングを行い、そこで議論した内容を次の練習に活用し、トレーニング効果を検証している。我々の提案方式が実現すれば、様々な視点からサッカーの試合を観察可能な Bullet-Time 映像が生成されるため、試合の状況の深く把握することが可能である。

3. 動的シーンに対応した Bullet-Time 映像生成

図1に、本研究で提案する動的シーンに対応した Bullet-Time 映像の自動切替手法の処理を示す。

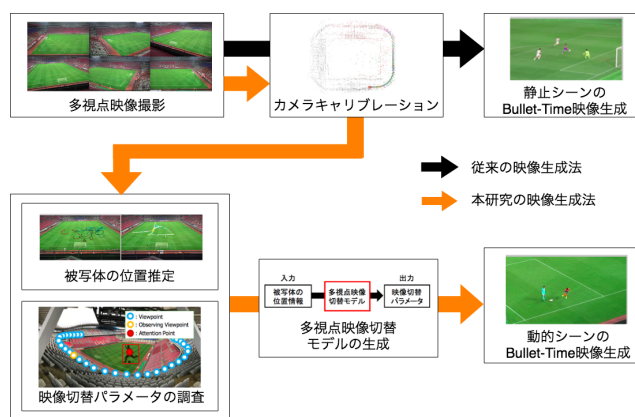


図1 動的シーンに対応した Bullet-Time 映像の映像切替パラメータ推定の流れ

サッカースタジアムを複数台のカメラで撮影し、多視点

映像を取得する。多視点映像を撮影したカメラのカメラパラメータをカメラキャリブレーションによって推定する。推定したカメラパラメータを参照することで、従来手法のような静止シーンの Bullet-Time 映像生成が可能となる。多視点映像に対して物体追跡処理を施し、被写体（選手やボール）の位置情報を推定する。観察者に撮影シーンの状況把握に適した映像切替パラメータに関する調査を実施する。選手、ボール、ゴールの位置情報と観察者調査で収集した映像切替パラメータの関係をニューラルネットワークで学習（深層学習）し、位置情報に基づいた多視点映像切替モデルを生成する。映像閲覧時には、観察シーンの選手、ボール、ゴールの位置を多視点映像切替モデルに与え、出力された映像切替パラメータを用いて多視点映像を切り替えることにより、動的シーンに適用した Bullet-Time 映像生成を実現する。

4. 多視点映像情報の取得

4.1 多視点映像の撮影

サッカースタジアムにおける撮影の一例を図 2 に示す。ここでは、図 2 の赤枠で囲まれる観察したい被写体が存在する場所を取り囲むようにカメラを配置する。サッカースタジアムなどの大規模空間で撮影する場合には、スタジアムの手摺などにカメラを固定し、図 2 の黄線のように観察したいエリアの中心で全てのカメラの光軸が交わるように向きを揃える。カメラの配置場所は、隣り合うカメラの光軸間のなす角がほぼ同じ角度になるように配置するのが好ましい。滑らかな視点切替を再現可能な Bullet-Time 映像を提示するためには、撮影用カメラを密に配置する必要がある。カメラの台数を増やし、隣り合うカメラの光軸のなす角をより狭くすると、より滑らかに視点が移動することができる。

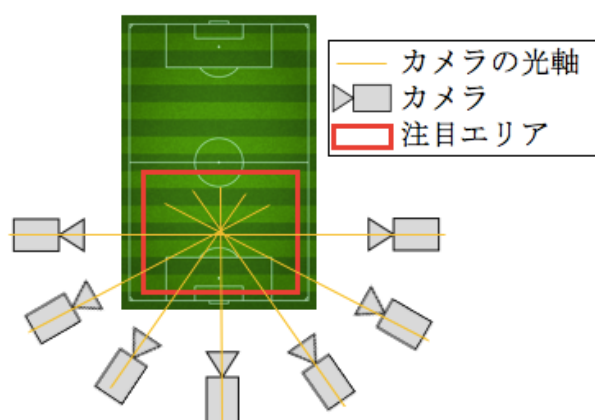


図 2 大規模空間での多視点映像の撮影

4.2 カメラキャリブレーション

Bullet-Time 映像を生成するためには、多視点カメラの位置姿勢、焦点距離などのカメラパラメータが必要である。フィールド上での選手とボールの位置情報を推定する際にもカメラパラメータが必要となる。本研究では、Structure from Motion (SfM)[17,18]を利用し、多視点画像間の対応点情報から SfM 座標系でのカメラの位置姿勢、焦点距離を推定する。SfM で生成された点群情報から、サッカーフィールドのコーナーやゴールポストなどランドマークとなる 3 次元点を検出し、ワールド座標系との剛体変換を求める。

5. 選手位置推定

5.1 背景差分法による前景領域分割

撮影した多視点映像に対して背景差分処理を適用し、画像の各画素について“前景領域らしさ”を表す前景領域度を算出する。背景画像は、時系列映像から一定時間区間で取り出した画像の平均画像として逐次生成する。

5.2 ボクセル空間の設定と前景領域度の投票

前節で求めた前景領域度を用いて、サッカーフィールド上での選手の 3 次元位置を推定する。まず、図 3 に示すようにサッカーフィールド上にボクセル空間を設定する。

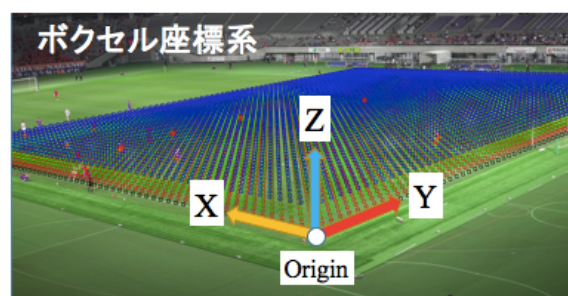


図 3 サッカーフィールド上でのボクセル空間

図 3 のように、ボクセル空間はフィールドのあるコーナーを原点、そのコーナーと接するタッチラインを X 軸、ゴールラインを Y 軸とし、Z 軸は X 軸 Y 軸と直交するように設定される（4.2 節で述べたワールド座標系と一致する）。また、ボクセルのサイズは、X 方向 Y 方向は観察するサッカーフィールドの広さによって、Z 方向は選手の身長に基づいて設定する。

全多視点画像で取得した前景領域度を、各カメラの射影変換行列を用いてボクセル空間に投影する。以降この投影処理を投票と呼ぶ。同一選手の前景領域度は、その選手が存在する 3 次元空間（フィールド）の同一ボクセル付近に集中して投票される。つまり、選手などの前景領域が存在するボクセルには高い前景領域度が投票され、存在しないボクセルに投票される前景領域度は低くなる。

5.3 選手の区別及び選手位置追跡

ボクセル空間で極大値探索を行う。前節で述べたように、検出された極大値を有するボクセルには、前景領域が存在している可能性が高い。その後、検出されたボクセル統合し選手位置を推定する。検出された前景領域の周辺を探索し、近くに前景領域が存在する場合は、同一選手の情報として統合する。統合した領域の体積（ボクセル数）が一定閾値よりも小さい場合、その領域は選手以外の物体と判断する。この際の探索範囲は、例えば $1\text{m} \times 1\text{m} \times 2\text{m}$ など一般的な体格の人が十分に収まる範囲で行う。選手領域を統合した結果から注目フレームの選手位置を推定する。

一つ前のフレームで検出された位置情報と現フレームの関連付けを行うことにより、選手軌跡を推定する。具体的には、現フレームと一つ前のフレームでの推定位置が、一定距離よりも短い場合、同一選手として関連付ける。この距離は一般的なサッカー選手の走行速度と撮影カメラのフレームレートによって規定する。

6. ボール位置推定

6.1 ボール候補位置の検出

多視点画像中で観測されるボール位置からその3次元位置情報を推定する。図4に示すボールテンプレート画像を用いて各画像におけるボール候補位置を検出する。

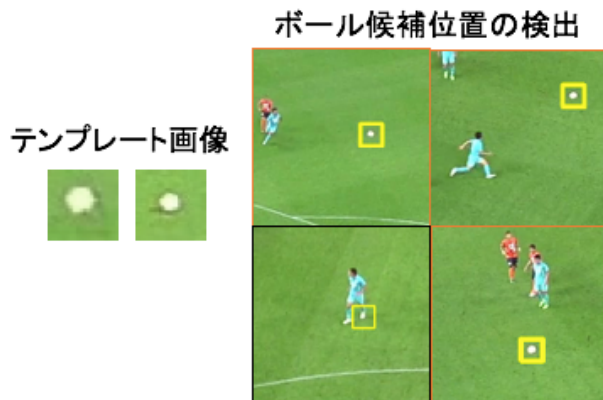


図4 ボール候補位置の検出

ボールテンプレート画像は、実際に撮影した画像からボールの3次元位置に応じて複数枚準備し、ボールの観測サイズの変化に対応する。撮影画像とテンプレート画像の類似度を計算し、類似度が閾値以上の領域をボール候補位置とする。この結果には、図4右の黄色枠に示すように、シューズなどボール以外の領域も含まれるが、後段の処理でフィルタリングするため、真のボール候補位置を検出漏れしないよう閾値を設定する。

6.2 エピポーラ線を用いたボール候補位置の限定

図4右の画像では、選手の白いスパイクをボールとして誤検出している。複数画像間でのエピポーラ方程式に基づいた幾何学的拘束により、この誤検出をフィルタリングする。本処理は、多数のカメラを用いた様々な方向からの撮影が行われていることを前提とする。

サッカーボールは球体であるため、どのような角度から観察してもボールテンプレート画像と同じ円形領域物体として観察される、つまり、正しいボール領域は、多くの視点において候補位置として検出されやすい。一方で、シューズの形状は球体とは異なるため、非円形領域として観測される視点では候補位置として検出されにくい。

各カメラの射影変換行列に基づいてカメラ間のエピポーラ方程式を算出する。ある視点画像で検出したボール候補位置に対応する他の視点画像におけるエピポーラ線を求め、その線上でのボール候補位置の存在を確認する。図5に、ボール候補位置を探索している様子を示す。図5のように、正しいボール候補位置を活用したエピポーラ線上には、複数視点でボール候補位置が検出される。

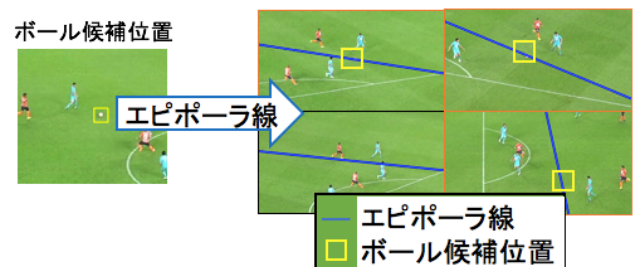


図5 エピポーラ線上でのボール候補位置の探索

6.3 多眼ステレオによる3次元位置推定

多眼ステレオによりボールの3次元位置を推定する。前節の処理によってボール候補位置とされた領域が2視点以上で存在する場合、全てボール候補位置でペアを組み、多眼ステレオによって3次元位置を推定する。多眼ステレオに使用するペアを複数作ることが可能な場合は、推定結果の外れ値の影響を避けるために、全推定結果の中央値をボールの3次元位置とする。

6.4 推定に失敗したボール位置の補間処理

ボールが選手の足元に存在するなど、ボール候補位置が検出されなかった場合、前後のフレームで推定されたボールの位置から注目フレームのボール位置を補間する。具体的には、XY平面では等速直線運動モデルを、Z方向は重力加速度による等加速度直線運動モデルを用いて補間する。

7. 位置情報を活用した多視点映像切替

本章では、サッカー試合の状況を表す重要な情報源である選手、ボール、ゴールの位置情報に基づき、映像切替パラメータを推定する手法について述べる。

7.1 Bullet-Time 映像生成に必要な映像切替パラメータ

Bullet-Time 映像生成に必要な要素として、何処に注目するか（注視点）、どの方向から観察するか（観察視点）、どのくらいの大きさで観察したいか（ズーム値）の設定がある。図 6 に、三つの映像切替パラメータを用いて Bullet-Time 映像生成の処理を説明した図を示す。まず、多視点映像から、観察に好ましい映像を一つ選択し、観察視点を設定する。次に、視点切替の際に中心となる点である注視点を設定する。注視点の位置は、図 6 の画面中央の赤丸で表示されており、注目したい物体に設定する。最後に、被写体の観察サイズを調整するためにズーム値を設定する。

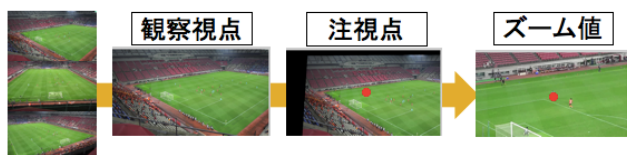


図 6 Bullet-Time 映像生成に必要なパラメータ
(映像切替パラメータ)

7.2 映像切替パラメータの調査

サッカーにおける状況把握に適した Bullet-Time 映像の映像切替は、複数の選手が広いフィールドを動き回るため、個人の嗜好差によって注目したい選手や観察したい視点が異なることが予想される。本節では、観察者毎にカスタマイズした映像切替パラメータの推定手法について述べる。

個々のユーザの好みを反映した多視点映像切替モデル生成するために、観察者毎に Bullet-Time 映像の映像切替パラメータを調査する。サッカーの試合では、攻撃と守備、味方と相手など注目要素によって注目要素が変化する可能性があるため、片方のチームがゴール前に迫っていく攻撃シーンなど、映像切替パラメータを取得するシーン毎に調査が必要となる。また、映像切替パラメータを取得する際には、サイドでの攻撃（左右）、中央での攻撃など注目するエリアの全体で選手がプレーしているシーンを活用する。

撮影したサッカー試合の全フレームにおいてパラメータ調査を行なうことは、膨大な労力と時間を要するため現実的でない。本研究では、一定時間間隔で多視点画像を準備し、そのフレームでの映像切替パラメータを調査する。切り出す時間間隔は、サッカー選手の一人当たりのボール所有時間を参考に決める。例えば、2007 年のドイツ代表の試合では、一人当たり 1.6 秒以内にボールを受け取ってから

パスを出している[19]。この情報を活用すると、1.6 秒以内の間隔で画像を取り出せば、切り出した画像でボール所有選手を確認することができる。ボール所有選手を確認できれば、試合の流れが把握しやすいため、シーンに適した映像切替パラメータが設定できる。

7.3 ニューラルネットワークを用いた多視点映像切替モデルの生成

ニューラルネットワークを用いた多視点映像切替モデル生成処理の流れを図 7 に示す。図 7 の上段の学習フェーズでは、ニューラルネットワークを用いて多視点映像切替推定モデルを生成する。学習に用いるデータとしては、位置情報と映像切替パラメータを使用する。位置情報としては、選手、ボール、ゴールの位置情報を与える。映像切替パラメータは、注視点、観察視点、ズーム値を与える。学習フェーズで使用する映像切替パラメータは、観察者調査によって取得したデータを与える。

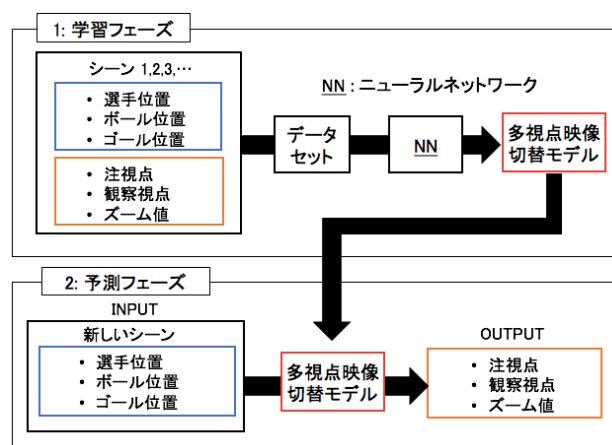


図 7 多視点映像切替モデルによる映像生成の流れ

位置情報と映像切替パラメータの関係性をニューラルネットワークで学習するために、各データを正規化したデータセットを準備する。選手、ボール、ゴールの位置情報は、サッカーフィールドの大きさを正規化する。注視点は、サッカーフィールド上の座標が格納されるため、サッカーフィールドの大きさを正規化する。視点は、カメラ台数で割ることで正規化を行う。ズーム値は取りうる値の最小値と最大値で正規化する。

図 7 の下段の予測フェーズでは、生成したモデルを使用し、映像切替パラメータを予測する。学習に使用していない新しいシーンの位置情報を入力データとして多視点映像切替モデルに入力し、出力データとして予測された映像切替パラメータを取得する。取得した映像切替パラメータを使用することで、動的シーンに適応した Bullet-Time 映像の自動生成が実現される。

8. 評価実験

8.1 撮影実験

本研究では、2017年9月20日にカシマサッカースタジアムで開催された第97回天皇杯全日本サッカー選手権大会の試合を撮影した映像を用いて実証実験を行った。図8にカメラの配置図を以下に示す。31台の多視点4Kカメラ(SONY FDR-AX100)で撮影を行い、30fpsの映像を撮影した。図8に示すように、全てのカメラは、片方のペナルティエリアのラインの中で、ゴールラインと平行であるラインの中央にカメラの画面中央を向くよう固定した。また、隣り合うカメラの光軸間のなす角が 6° になるようカメラを配置した。



図8 多視点カメラの配置図

8.2 映像処理環境

映像処理はCPU: Intel Core i7-4770 3.40GHz, メモリ: 16.0GBを装備したPCを用いて行った。Bullet-Time映像は、Akechiら[7]の手法を用いて生成する。カメラキャリブレーションは、一般公開SfMライブラリのVisualSFM[20]を利用した。

8.3 位置情報推定システムの実装

選手とボールの位置情報は、多視点カメラで重複撮影を行ったサッカーフィールドの半面において推定した。選手位置推定では、ボクセルの大きさをカシマスタジアムのフィールド半面の大きさ(57.5m × 78m) [21], 人間の身長より大きい高さ2mに設定した。また、ボクセルの間隔は0.2mに設定した。ボクセル値の極値を検出後、その極値周辺1m × 1m × 2mで別の極値が存在した場合、同じ選手情報として関連付けをした。また、現在のフレームで検出された選手位置のXY平面上での0.4m以内に一つ前のフレームで検出された選手位置が含まれる場合、同一選手の位置情報としてフレーム間で関連付けし、選手位置追跡を行なった。図9に各選手の追跡結果の一例を示す。図9では、各選手の移動軌跡を色分けし、高さZ=1mの位置情報を取り出して表示している。図9から、選手の位置情報を取得できていることが確認できる。



図9 選手の移動軌跡の一例

ボール位置推定では、テンプレート画像を2種類用意し、ボール位置候補位置の検出を行なった。また、テンプレート画像との類似度の評価には、正規化相互相関(NCC)で使い、NCCの値が0.7以上の値をボール候補位置として検出した。その後、エピポーラ線を活用した視点数の制約条件は、15台以上の視点でエピポーラ線上にボール候補位置が存在しなかった場合に、ボール候補位置から除外するように設定した。図10にボールの追跡結果の一例を示す。選手位置で補間した位置を黄色、線形補間した位置を水色で示している。図10から、ボールの位置情報を正しく推定できていることが確認できる。

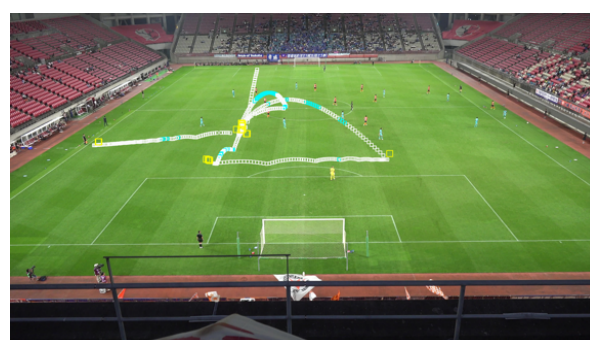


図10 ボールの移動軌跡の一例

8.4 ニューラルネットワークによる多視点映像切替モデルの実装

ニューラルネットワークの実装には、Google Colaboratory[23]を使用した。また、Keras[24]のNNライブラリを活用してニューラルネットワークの実装を行なった。また、モデル生成に用いるデータセットから訓練データを8割、テストデータを2割になるようランダムに分割した。訓練データとテストデータの分割処理をランダムに5回行い、五つのモデルを生成できるよう実装した。生成した五つのモデルから、映像切替パラメータの予測結果を五つ取得し、取得した五つの予測結果の平均値を予測結果として活用した。多視点映像切替モデルの生成は、被験者毎に行なった。



図 11 被験者の好む Bullet-Time 映像 (映像 1)



図 12 多視点映像切替モデルによって自動生成させた Bullet-Time 映像 (映像 2)

8.5 被験者実験

本研究では、7 人の被験者を対象に実験を行った。撮影した試合映像から、5 秒間のシーンを 5 シーン分取り出した。5 シーンから、1 シーンをパラメータ設定練習用、3 シーンをモデル生成用、残り 1 シーンをモデル評価用として活用した。5 秒間の映像から 1 秒ごとに取り出した多視点画像を使用し、被験者から映像切替パラメータ（注視点、観察視点、ズーム値）を 1 シーン 6 箇所取得した。図 11 と図 12 に、被験者の好む Bullet-Time 映像と多視点映像切替モデルが生成した Bullet-Time 映像の比較画像の一例を示す。図 11 の映像 1 は、事前に被験者から取得した映像切替パラメータを活用しており、被験者の好む映像である。そのため、モデルによって生成させた図 12 の映像 2 が映像 1 と同じであれば、状況把握に適した映像が自動生成できたことになる。また、試合状況を把握する際に適した映像は、1 種類の映像とは限らない。そのため、映像 2 が映像 1 と同じでない場合でも、映像 2 が被験者にとって状況把握に適した映像である可能性がある。

被験者による主観評価実験として、映像 1 と映像 2 を観察しながら、4 つの質問に答えてもらった。質問に対しては、5 段階評価（「5:できている」、「4:どちらかといえばできている」、「3:どちらともいえない」、「2:どちらかといえばできていない」、「1:できていない」）を行ってもらった。質問内容を以下に示す。

- Q1. 映像 2 の注視点は、映像 1 と同じ位置を注視することができるか
- Q2. 映像 2 の観察視点は、映像 1 と同じ観察視点を選んでいるか
- Q3. 映像 2 のズーム値は、映像 1 と同じズームを選んでいるか
- Q4. 映像 2 は、試合状況の観察に適した映像が生成できているか

図 13 に、上記四つの質問に対して 7 人の被験者が答え

た評価結果を示す。

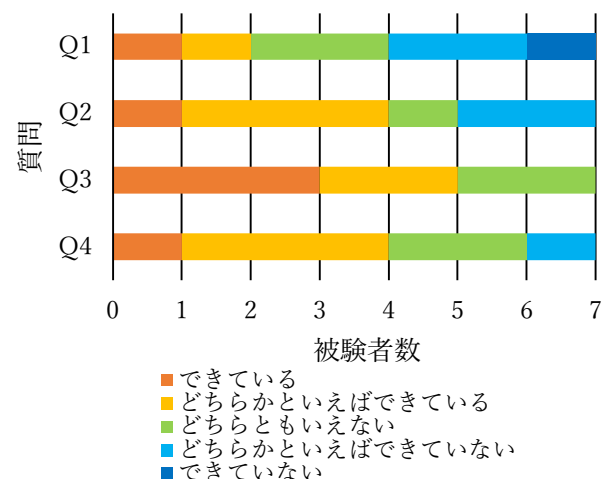


図 13 被験者による主観評価結果

図 13 のグラフでは、7 人の被験者による四つの質問に対する 5 段階評価の結果を示している。Q1, Q2 の評価結果は、注視点と観察視点を同じでないと評価した被験者が存在することが確認できる。そのため、被験者が好む注視点、観察視点を多視点映像切替モデルで再現できなかったケースが存在した。Q3 の評価結果は、半数以上がズーム値は同じであると評価しており、多視点映像切替モデルによって被験者の好むズーム値を再現できる傾向が確認できた。Q4 の評価結果は、半数以上が試合状況の観察に適した映像であると評価した結果が得られた。生成した多視点映像切替モデルを活用することで、試合状況の観察に適した映像が生成される傾向が確認できた。

これらの結果から、あらかじめ取得した被験者が好む注視点や観察視点は全てのケースで再現できたとは言いが、被験者が試合状況の観察に適した映像であると感じる映像を自動で生成できた傾向が確認できた。

9. おわりに

本研究では、動的シーンとしてサッカーの試合、大規模空間としてサッカースタジアム、多視点映像提示手法として Bullet-Time 映像を採用し、大規模空間での動的シーンに対応する多視点映像の自動切替手法を提案した。Bullet-Time 映像の自動生成を実現させるために、選手、ボール、ゴールの位置情報を活用し、ニューラルネットワークによる多視点映像切替モデルを生成した。多視点映像切替モデルによって予測された映像切替パラメータを活用することで、試合状況の観察に適した映像を生成できる傾向が確認できた。

参考文献

- [1] 稲本奈穂, 斎藤英雄, “視点位置の内挿に基づく 3 次元サッカー映像の自由視点観賞システム”, 映像情報メディア学会誌 Vol.58 No.4, pp.529~539, 2004.
- [2] 古山孝好, 向川康博, 亀田能成, 大田友一, “屋外大規模空間における自由視点映像生成のための選手領域抽出法”, 画像の認識・理解シンポジウム(MIRU2005), pp.1412-1419, 2005.
- [3] Jean-Yves Guillemaut and Adrian Hilton, “Joint Multi-Layer Segmentation and Reconstruction for Free-Viewpoint Video Applications”, International Journal of Computer Vision Volume 93 Issue 1, pp. 73-100, 2011.
- [4] Joel Carranza, Christian Theobalt, Marcus A. Magnor and Hans-Peter Seidel, “Free-viewpoint video of human actors”, ACM SIGGRAPH 2003 Papers, pp.1-9, 2003.
- [5] Aljoscha Smolic, Karsten Mueller, Philipp Merkle, Christoph Fehn, Peter Kauff, Peter Eisert and Thomas Wiegand, “3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards”, IEEE International Conference on Multimedia and Expo, pp.2161-2164, 2006.
- [6] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk and Steve Sullivan “High-quality streamable free-viewpoint video”, ACM Transactions on Graphics (TOG) Volume 34 Issue 4, 2015.
- [7] Nao Akechi, Itaru Kitahara, Ryuuki Sakamoto and Yuichi Ohta, “Multi-Resolution Bullet-Time Effect”, In ACM SIGGRAPH-Asia, 2014.
- [8] Kensuke Ikeya and Yuichi Iwade, “Multi-viewpoint robotic cameras and their applications”, ITE Transactions on Media Technology and Applications Volume 4 Issue 4, pp.349-362, 2016.
- [9] Christine Chen, Oliver Wang, Simon Heinzle, Peter Carr, Aljoscha Smolic and Markus Gross, “Computational Sports Broadcasting: Automated Director Assistance for Live Sports”, IEEE International Conference on Multimedia and Expo (ICME), 2013.
- [10] Hao Jiang, Sidney Fels and James J. Little, “Optimizing Multiple Object Tracking and Best View Video Synthesis”, IEEE Transactions on Multimedia (Volume: 10 , Issue: 6), pp. 997-1012, 2008.
- [11] Sachiko Iwase and Hideo Saito, “Parallel Tracking of All Soccer Players by Integrating Detected Positions in Multiple View Images”, Proceedings of the 17th International Conference on Pattern Recognition, 2004.
- [12] 糟谷望, 北原格, 亀田能成, 大田友一, “サッカーシーンにおける選手視点映像提示のためのリアルタイム選手軌跡獲得手法”, 画像電子学会誌 38(4), pp.395-403, 2009.
- [13] Jinchang Ren, James Orwell , Graeme A. Jones and Ming Xu, “Tracking the soccer ball using multiple fixed cameras”, Computer Vision and Image Understanding 113, pp.633-642, 2009.
- [14] 石井規弘, 北原格, 亀田能成, 大田友一, “2 視点からの映像を利用したサッカーボール追跡手法”, MIRU2008 画像の認識・理解シンポジウム 8, pp.1035-1040, 2008.
- [15] Xinguo Yu and Dirk Farin , “Current and Emerging Topics in Sports Video Processing”, IEEE International Conference on Multimedia and Expo, pp.1-4, 2015
- [16] 樋口智洋, 堀野博幸, 土屋純, “大学サッカーにおける戦術トレーニング効果の検討 - 「プレー重心」を用いて - ” , スポーツパフォーマンス研究 5, pp.176-188, 2013.
- [17] Carlo Tomasi and Takeo Kanade, “Shape and motion from image streams under orthography: a factorization method”, International Journal of Computer Vision 9-2, pp.137-154, 1992.
- [18] Changchang Wu, “Towards Linear-time Incremental Structure from Motion”, International Conference on 3D Vision, pp.127-134, 2013.
- [19] 木崎伸也, “サッカーの見方は一日で変えられる”, pp62-66, 東洋経済新報社, 2010.
- [20] Changchang Wu, “VisualSFM: A Visual Structure from Motion System”, <http://ccwu.me/vsfm/>, (参照 2018/12/27).
- [21] “施設概要 | 茨城県立カシマサッカースタジアム”, <http://www.so-net.ne.jp/antlers/kashima-stadium/about/>, (参照 2018/12/27).
- [22] J J Hernandez Gomez, V Marquina and R W Gomez, “On the performance of Usain Bolt in the 100 m sprint”, in European Journal of Physics 34(5), pp. 1227-1233, 2013.
- [23] “こんにちは Colaboratory - Colaboratory”, <https://colab.research.google.com/notebooks/welcome.ipynb>, (参照 2018/12/27).
- [24] “Home - Keras Documentation”, <https://keras.io/>, (参照 2018/12/27).