

多視点映像を用いた簡易な操作による3次元背景モデル構築法

野村 尚史[†] 北原 格[‡] 亀田 能成[‡] 大田 友一[‡]

[†] [‡] 筑波大学大学院システム情報工学研究科 〒305-8577 茨城県つくば市天王台 1-1-1

E-mail: [‡] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

あらまし スポーツシーンにおける自由視点映像生成技法の一環として、オペレータの簡易な操作により自由視点映像の背景領域を生成する手法を提案する。本手法は、スポーツフィールドのような大規模空間においては、背景が競技スペースに対して比較的遠くに存在しているという特性を利用している。フィールドを取り囲むように設置した複数のカメラ映像に対し、オペレータの操作としては、背景領域上の数点を指定するだけよい。その点情報を元に、システムは簡易背景3次元モデルを構築し、映像を仮想的に背景を正面から撮影したように変形し3次元モデルにテクスチャとして貼り付けることにより、つなぎ目にズレのない、自然な見え方の自由視点映像生成を実現する。スタジアムや体育館などの大規模空間において撮影した多視点映像を用いて背景の自由視点映像の生成実験を行い、提案手法の有用性を確認した。

キーワード 自由視点映像, 3次元モデリング, テクスチャマッピング, スポーツシーン, 平面近似

3D Background Modeling Method with Simple Operations for Multi-View Videos

Naobumi Nomura[†] Itaru Kitahara[‡] Yoshinari Kameda[‡] and Yuichi Ohta[‡]

[†] [‡] Graduate School of Systems and Information Engineering, University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573, Japan.

E-mail: [‡] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

Abstract This paper proposes a novel 3D background modeling method that only needs simple manual operations in generating a free-viewpoint video for sport scenes. To generate a 3D background model, background area in an image is divided into “distant region” and “ground region”. The 3D model of distant region is represented as textured vertical rectangular planes. The ground region is modeled as a simple horizontal plane. These two regions are segmented based on a few points given by manual operation. In order to show the validity of our method, we generated a free-viewpoint video of 3D background scene from multiple images recorded in a stadium and a gymnasium.

Keyword Free Viewpoint Video, 3D modeling, Texture Mapping, Sports Scene, Plane Fitting

1. はじめに

近年、コンピュータ技術の発展にともない、複数台のカメラを用いてさまざまな角度から映像を獲得し、それらの映像を計算機内部で融合することにより、自由な視点から映像を閲覧する自由視点映像メディアに関する研究が活発に行われている[1][2]。自由視点映像メディアは、娯楽、研究、教育、文化財保護などのさまざまな分野において応用が期待されているが、我々は特に、サッカースタジアムなどの大規模空間を撮影対象とし、多視点映像の撮影から3次元映像処理、ネットワークを介したデータ伝送、自由視点映像提示までを実時間で行う、自由視点映像中継方式に注目し、研究開発を行っている[3][9]。

その中でも、特に本稿では背景領域の映像生成手法について述べる。スポーツが行われているフィールドやプレーヤの可視化に比べて、背景部分について取り

上げている研究は少ない。しかし、臨場感のあるスポーツ自由視点映像生成のためには背景も重要である。現在、背景の構築方法として最もよく行われているのは背景部分をあらかじめ3次元モデルとして用意しておくことであるが、あらかじめ3次元モデルを用意するのは状況によっては簡単ではなく、可用性の高い方法とは言えない。

対象がプレーヤであるか背景であるかを問わず、自由視点映像における従来研究は、物体の3次元形状を陽に復元しない手法と復元する手法とに大別することができる。前者の例として、Zitnickらは、8台のカメラにより撮影された映像を Segmentation based stereo algorithm を用いて解析することにより、カメラ間の中間視点映像を生成する手法を提案している[4]。しかしこのような3次元形状を陽に復元しない手法では、多視点画像間での対応点探索という計算コストの高い処理が必要となるため実時間化が困難であるという問題

や、多視点カメラを設置した位置を内挿した箇所にか仮想カメラを設置できないという問題が存在するため、広い範囲を移動するスポーツ向けの自由視点映像の背景生成には不向きである。

これに対し3次元を陽に復元する手法は、視点設定の自由度が高いので、スポーツ向けの自由視点映像の背景生成に向いているといえる。Kanadeらは、半径5mの半球内に設置された51台のカメラにより撮影された多視点映像をもとに人物の正確な3次元形状を復元し、自由視点映像生成手法を提案している[5]。この手法は計算量がかかり、データ量も必要となるため、大規模空間への対応やリアルタイムでの映像生成への適応が困難である。これに対して、Koyamaらは被写体となる人物(前景領域)の3次元形状を1枚の平面形状(ビルボード)で近似することにより、撮影から提示までの処理を実時間で実行可能とする手法を提案し、サッカースタジアムのような大規模空間における自由視点映像生成・伝送・提示を実現している[3]。しかし、これらの従来研究では、シーン中の物体や人物といった前景領域に注目したものがほとんどであり、背景部分のように大きな広がりを持った対象には不向きである。

そこで本稿では、3次元モデルが用意できないような運動施設においても高品質な背景領域の提示を伴う自由視点映像を生成することを目的とし、オペレータの簡易な操作のみにより背景の3次元モデルを生成する手法について提案する。なお、背景モデルの生成方法は幾何的要素とテクスチャの要素から構成されるが、そのうち、幾何的要素の計算方法については[9]でも述べられている。

2. 簡易操作による背景生成手法

大規模空間を撮影対象とした場合、背景は遠方に存在している遠景領域とそれ以外の地面領域に二分することができる。本節では、各領域に対して適切な処理を施すことにより、自由視点映像を生成する手法について述べる。

2.1.背景領域の3次元モデル

本手法では、スタジアム、体育館などに設置された、カメラパラメータが既知の複数台のカメラで撮影された多視点映像を入力とする。

我々は大規模空間におけるスポーツイベントでは、以下のような特性が背景に当てはまると考え、カメラにより撮影された背景画像を、コート、ピッチといった地面領域と、壁・観客席といった遠景領域に二分する[9]。

- 選手(前景領域)はコート、ピッチといった1枚

の平面上(地面領域)で競技を行う。

- 背景領域は主に地面領域と壁・観客席のような遠景領域によって構成されている。
- 遠景領域は、カメラから十分に遠いため、地面領域に対して直立する複数の平面と近似することができる。
- 撮影に用いるカメラは地面領域に垂直に設置されている

一般に観客者は、ピッチ、コート上でプレーする選手に注目することが多く、壁や観客席などの遠景領域に注目することは少ない。また遠景領域はカメラから十分に遠方に存在しているため、3次元的な見え方の変化を正確に再現する必要は少ないと考えられる。我々は、これら遠景領域が持つ特性に着目し、複数枚の平面群として形状を近似し、それらに撮影映像のテクスチャを貼り付けることにより3次元モデルを構築する手法を提案する。映像上の数点を指定するという非常に簡易な処理だけで、背景領域を構成する平面群の分割処理を実現できる点が、本手法の大きな特長である。

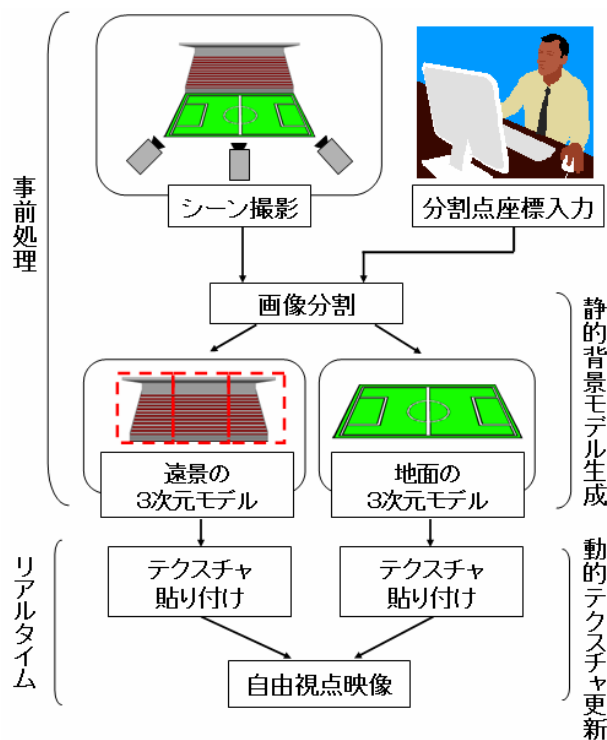


図1: 処理の概要

2.2.背景生成手法の概要

図1に、背景領域の自由視点映像生成処理の概要を示す。まず、撮影された多視点画像を地面と複数の遠景平面に分割する。この分割では、多視点画像中の地面領域と背景領域との境界を示す特徴点を手動で指定する必要があるが、1台の撮影カメラに対して、数点

のクリック処理を1度だけ行えばよく、それほど労力を要さない。入力多視点画像中の地面領域を2次元射影変換により3次元空間中の地面に投影し、それらを合成することにより地面の3次元モデルが構築される。さらに分割された小領域を3次元空間中に射影することにより、地面領域に直立する平面として、遠景の3次元モデルが構築される。最後に、3次元モデルに多視点映像のテクスチャを貼り付け、背景領域の自由視点映像を生成する。

提案手法は、分割毎に画像上の2点を指定するだけ、という簡易な操作により、自由視点映像生成が可能であるため、カメラ台数が増加した場合でも比較的容易に対応可能である。また、その分割間隔はユーザが生成結果を見ながらインタラクティブに操作できるため、労力に対して良好な生成結果が得られるという利点を有する。

なお、本手法ではいずれの撮影カメラも事前にキャリブレーションはされているものとする。自由視点映像生成の場合、プレーヤやフィールドの可視化においてもキャリブレーションは必須であるので、この前提は新たな作業量を要求するものではない点に注意する。

3. 地面モデルの構築

3.1. 地面領域モデルの生成

本節では、映像から地面の3次元モデルを構築するために必要な地面と画像平面間の座標変換について述べる[7][8]。3次元空間中のある点 (x,y,z) と、画像上での観測点 (u,v) との間の幾何的關係は、撮影に用いたカメラの射影変換行列を用いて、式(1)のように表すことができる。

$$\begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (1)$$

ただし、モデル復元処理を簡易化するため、 z 軸は鉛直上向きにとるものとする。地面を高さ $z=h$ の平面とすると、式(1)は次式のように書き直すことができる。

$$\begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & hC_{13} + C_{14} \\ C_{21} & C_{22} & hC_{23} + C_{24} \\ C_{31} & C_{32} & hC_{33} + C_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = B \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (2)$$

式(2)右辺の 3×3 行列 B は明らかに逆行列が存在するので、この逆行列 B^{-1} を求めることにより、画像上の点 (u,v) を地面領域上の任意の点 (x,y,h) に投影することが可能となる。この逆射影変換行列 B^{-1} を用いて、

撮影画像を3次元空間中の地面に投影し、地面モデルを構築する。

この幾何モデル上にさらに撮影視点ごとに構築された地面モデルのテクスチャを合成することで最終的な地面モデルを構築する。図2に10台のカメラからの地面テクスチャを均等重み付けで合成した例を示す。

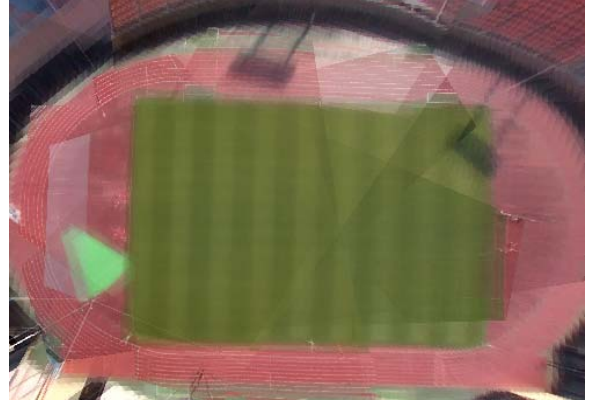


図2：合成処理により構築された地面モデル

3.2. 仮想視線の向きにもとづいたテクスチャの合成

撮影視点ごとに構築された地面モデルのテクスチャを均等重み付けで合成した場合、ポールや人物などの影が放射状に見えてしまうという問題や、逆光や光の照り返しなどが正しく表現できないという問題が存在する。そこで仮想視線の向きに応じて最適な地面テクスチャを選択し、描画することでより臨場感のある自由視点映像を生成する手法を提案する。

仮想視線の x - y 平面上での向きを θ 、カメラの向きを θ_m ($m=1,2,\dots,n$)とする。地面領域を囲むようにカメラ $1 \sim n$ が設置されていると仮定すれば、地面付近での仮想視線に対して地面を同じ方向から映しているカメラ \hat{m} は式(3)によって得られる。

$$\hat{m} = \arg \min_m |\theta - \theta_m| \quad (3)$$

このカメラ \hat{m} から得られるテクスチャを、3.1節にて述べた合成地面テクスチャの上に重ねて描画することで地面領域の自由視点映像を生成する。

4. 遠景モデルの構築

4.1. 分割画像群による遠景領域モデルの生成

本手法では、遠景領域を実形状とは異なる複数の平面で表現する。図3に示すように、遠景領域を小領域に分割し、その分割小領域ごとに3次元平面近似を行うことにより近似精度を高め、歪みの少ない背景自由

視点映像を生成する[9].

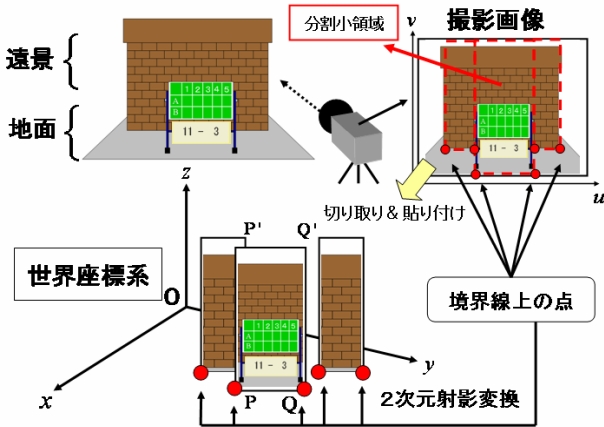


図3：遠景領域分割処理による形状復元

まず、入力画像中の遠景領域と地面領域の境界線上2点の座標 $\mathbf{p} = (u_p, v_p)$, $\mathbf{q} = (u_q, v_q)$ を手動で指定し、その2点を下端とする矩形領域を分割小領域として切り取る。この2点を以降、分割点と呼ぶ。次に2つの分割点 \mathbf{p} , \mathbf{q} を式(2)を用いて地面上に射影し、3次元座標 $\mathbf{P} = (x_p, y_p, h)$, $\mathbf{Q} = (x_q, y_q, h)$ を算出する。3次元空間中に投影した2点 \mathbf{PQ} 間の距離 d と、分割画像のアスペクト比 α を式(4)に与えることにより、分割画像の上端隅の3次元空間中における対応点 \mathbf{P}' , \mathbf{Q}' を求めることができる (アスペクト比は分割小領域の高さを幅で割った値とする)。これら分割小領域の四隅の3次元座標から地面に直立する遠景モデルが構築される。

$$\begin{aligned} \mathbf{P}' &= (x_p, y_p, h + d\alpha) \\ \mathbf{Q}' &= (x_q, y_q, h + d\alpha) \end{aligned} \quad (4)$$

こうして事前に生成した遠景領域の遠景モデルに競技中に撮影した背景画像を貼り付けていくことにより自由視点映像を生成する。この手法は簡単でありながら背景領域を構築できるという点で有用である。しかし、この方法で正しく遠景領域をモデル化できるのは、撮影カメラが背景領域を構築する物体群に対して正対しているという制約が成り立つときである。

4.2. 分割小領域の2次元射影変換

複数の分割小領域を背景3次元形状モデルに貼り付けた場合、画像間の解像度の違い、画像中の壁の奥行きの違いなどが原因となり、図5(上)の入力に対して、図5(中)に示すように分割小領域のつなぎ目においてズレが発生する場合がある。このズレを解消するためには、遠景領域面を正面から撮影した映像を貼り付ける必要があるが、物理的制約から常にそのよ

うな撮影が行うことができるとは限らない。そこで本手法では、遠景領域の直立仮定を用いて、4点の対応関係から遠景領域平面と撮影画像間の2次元射影行列を求め、この2次元射影変換を用いて遠景領域を仮想的に正面から撮影した映像を生成することにより、分割小領域のつなぎ目におけるズレを解消する。

4.2.1. 2次元射影行列の算出

2次元射影変換は、次式で表される平面間の射影関係である。

$$\tilde{\mathbf{m}} = \mathbf{H}\tilde{\mathbf{m}}' \quad (5)$$

ここで $\tilde{\mathbf{m}}$, $\tilde{\mathbf{m}}'$ は2枚の平面間における対応点の同次座標である。4点以上の対応関係がわかっているならば、2次元射影行列 \mathbf{H} を求めることができる[6]。推定誤差を低減させるためには、4点以上の対応関係から最小自乗法を用いて2次元射影行列を算出するのが一般的であるが、本研究では操作の簡易性を優先し、4点の対応関係のみを用いて2次元射影行列を算出する。

4.2.2. 仮想的に正面から撮影した見え方の生成

既知である2つの分割点を点 \mathbf{p} , \mathbf{q} とする。次に式(2)を用いて点 \mathbf{p} , \mathbf{q} を3次元空間中の地面平面に射影し、この点をそれぞれ点 \mathbf{P} , \mathbf{Q} とする。点 \mathbf{PQ} 間の距離 d を求め、この2点の鉛直上方、距離 d のところに点 \mathbf{P}'' , \mathbf{Q}'' を定める。図4に示すように点 \mathbf{P} , \mathbf{Q} , \mathbf{P}'' , \mathbf{Q}'' は3次元空間中で正方形を構成する。式(1)により点 \mathbf{P}'' , \mathbf{Q}'' を撮影画像上に射影し、点 \mathbf{p}'' , \mathbf{q}'' を求める。4点 \mathbf{P} , \mathbf{Q} , \mathbf{P}'' , \mathbf{Q}'' が正方形に見えるように仮想正面カメラを用意し、このカメラから見た4点をそれぞれ $\hat{\mathbf{p}}$, $\hat{\mathbf{q}}$, $\hat{\mathbf{p}}''$, $\hat{\mathbf{q}}''$ とする。点 \mathbf{p} , $\hat{\mathbf{p}}$, 点 \mathbf{q} , $\hat{\mathbf{q}}$, 点 \mathbf{p}'' , $\hat{\mathbf{p}}''$, 点 \mathbf{q}'' , $\hat{\mathbf{q}}''$ の4組の対応関係から2次元射影行列 \mathbf{H} を算出し、この射影変換により、背景領域を正面から撮影したような見え方に撮影画像を変換し、3次元モデルに貼り付ける。各点の関係を以下に示す。

$$\begin{cases} \mathbf{p} = (u_p, v_p) \\ \mathbf{q} = (u_q, v_q) \\ \mathbf{p}'' = (u_p'', v_p'') = \mathbf{B}^{-1}\mathbf{P}'' \\ \mathbf{q}'' = (u_q'', v_q'') = \mathbf{B}^{-1}\mathbf{Q}'' \end{cases} \quad (6)$$

$$\begin{cases} \mathbf{P} = (x_p, y_p, h) \\ \mathbf{Q} = (x_q, y_q, h) \\ \mathbf{P}'' = (x_p, y_p, h + d) \\ \mathbf{Q}'' = (x_q, y_q, h + d) \end{cases}$$

$$\begin{cases} \hat{\mathbf{p}} = (u_{\hat{p}}, v_{\hat{p}}) \\ \hat{\mathbf{q}} = (u_{\hat{q}}, v_{\hat{q}}) \\ \hat{\mathbf{p}}'' = (u_{\hat{p}}'', v_{\hat{p}}'' + (u_q - u_p)) \\ \hat{\mathbf{q}}'' = (u_{\hat{q}}'', v_{\hat{q}}'' + (u_q - u_p)) \end{cases}$$

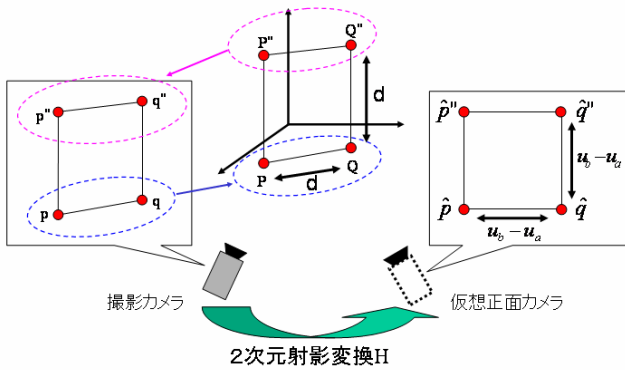


図4：撮影画像と形状平面間の2次元射影変換の推定

図5（上）に示した映像と分割点を与えた場合における2次元射影変換後の自由視点映像を図5（下）に示す。このように正面から背景領域を撮影していない映像でも、2次元射影変換を行うことで、分割小領域のつなぎ目におけるズレが解消されていることがわかる。

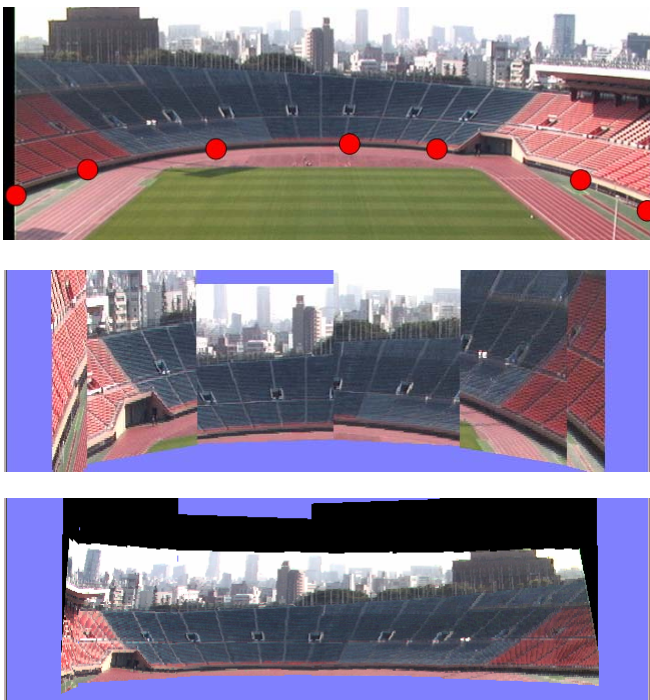


図5：背景領域の自由視点映像生成の例（上）入力画像と分割点の位置（中）撮影テクスチャをそのまま貼り付けた例、（下）正面から撮影

したテクスチャに変換後貼り付けた例

4.3. 背景画像の更新

サッカーの試合では、観客の入退場や、応援の盛り上がりによって背景が動的に変化する場合がある。そこで、本研究では、遠景領域にマッピングする背景画像を随時更新し、より臨場感のある自由視点映像を生成する。

5. 実験

図6に示すように、国立霞ヶ丘競技場に3CCDビデオカメラ(SONY, DXC9000)10台を設置して多視点映像の撮影実験を行った。撮影映像の解像度は640×480画像、フレームレートは毎秒30枚である。

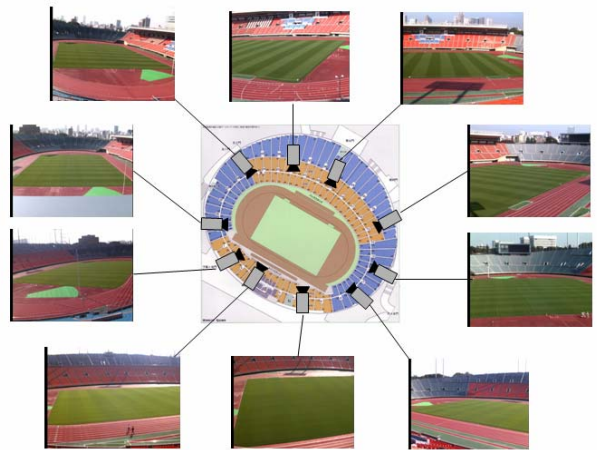


図6：霞ヶ丘競技場実験でのカメラ配置

入力多視点画像に対して、計34個の分割点を指定し17枚の分割小領域を取得し、背景領域の3次元モデルを生成した。入力分割点34点を平面図上に示したものが図7である。画像の分割に要した入力手作業の時間は10分程度であり、十分簡易であると言える。実際に生成された自由視点映像を図8に示す。地面と同じ高さに視点を設定した場合の見え方や、競技場上空から全体を俯瞰したような見え方など、自由な視点からの映像を観察することができる。地面モデルに貼り付けるテクスチャを仮想視点位置に応じて最適なテクスチャを選択し、重ねて描画することで、逆光や光の照り返しなどの様子が表現できていることがわかる。また、分割小領域のテクスチャを2次元射影変換し3次元モデルに貼り付けることで、つなぎ目のズレが目立たない自然な見え方が再現されていることがわかる。

国立スポーツ科学センター体育館においても、図6と同様にコートの周囲にカメラ(SONY,BRC300)を10台配置した撮影実験を行った。この実験で撮影した映像を用いて、4.3節で述べた背景画像の更新処理を行

った。図9に、人物が壁際を歩く様子を撮影した映像を用いて、背景領域の自由視点映像を生成した例を示す。背景の時系列変化に合わせて画像を更新することで、撮影空間の雰囲気により把握し易い自由視点映像が生成・提示されていることがわかる。

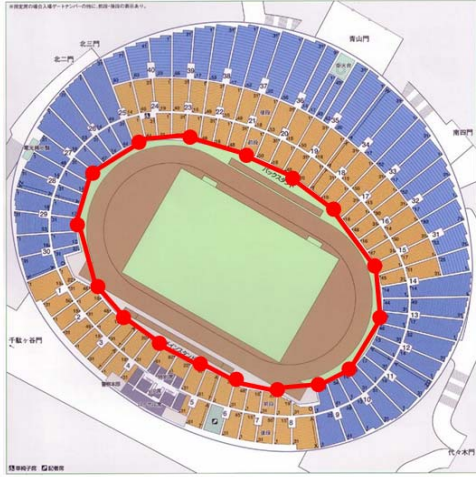


図7：霞ヶ丘競技場での分割点の空間的位置

6. まとめ

スポーツシーンを対象として、簡易な操作により背景領域における自由視点映像を生成する手法を提案した。本研究では、カメラから撮影された背景画像を地面領域と遠景領域に二分し、領域毎に自由視点映像を生成している。地面領域においては、背景画像中の地面領域を3次元空間中の地面に投影し、テクスチャを合成することで3次元モデルを構築した。また遠景領域においては、背景画像上の数点を指定するだけという簡易な操作により、写り込んだ遠景領域の3次元形状を適切に表現した3次元モデルを生成した。競技場や体育館など、実際の大規模空間において撮影した多視点映像を用いて自由視点映像を生成することにより、提案手法の有用性を確認した。今後は、映像の品質を

向上させるために、カメラ間の色合いを統一することなどが課題として挙げられる。

なお、本研究の一部は総務省戦略的情報通信研究開発推進制度“ネットワークによる自由視点映像のライブ配信とインタラクティブ提示”により実施した。

文 献

- [1] F. Caillette and T. Howard, “Real-time Markerless Human Body Tracking with Multi-view 3-D Voxel Reconstruction,” Proc. SIGGRAPH, pp.43-54, 1996.
- [2] S. Yamazaki, R. Sagawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi, “Microfacet Billboarding,” Proc. the 13th Eurographics Workshop on Rendering, pp.175-186, 2002.
- [3] T. Koyama, I. Kitahara, and Y.Ohta, “Live Mixed-Reality 3D Video in Soccer Stadium,” Proc. of the 2nd IEEE and ACM Int. Symposium on Mixed and Augmented Reality (ISMAR2003), pp.167-178, 2003.
- [4] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality Video View Interpolation Using a Layered Representation,” ACM SIGGRAPH and ACM Trans. on Graphics, Vol23, Issue 3, pp.600-608, 2004.
- [5] T. Kanade, P. Rander, and P. J. Narayanan, “Virtualized Reality: Constructing Virtual Worlds from Real Scenes,” IEEE Multimedia, Vol.4, No1, pp.34-37, 1997.
- [6] 徐剛, 辻三郎, “3次元ビジョン,” 共立出版, 1998.
- [7] R. Szeliski, “Video Mosaics for Virtual Environments,” IEEE Computer Graphics and Applications(c), Vol. 16, No.2, pp.22-30, 1996.
- [8] 市原栄太郎, 高尾広行, 大田友一, “NaviView: 仮想車載カメラ映像による運転者の視覚支援”, 電子情報通信学会論文誌, Vol.J82-D- II, No.10, pp.1816-1825, 1999.
- [9] 野村尚史, 北原格, 亀田能成, 大田友一, “自由視点映像生成のための簡易な操作による背景生成手法”, 電子情報通信学会技術研究報 PRMU, Vol.103, No.429, PRMU2006-176, pp.31-36, 2006.



図8：生成した自由視点映像の例



図9：背景画像を更新した例