I-030

レンダリング時の背景モデル参照を用いたリアルタイム自由視点映像の高品質化に関する検討
Study on visual quality improvement of real-time free viewpoint video using background model in rendering

渡邊 良亮† Ryosuke Watanabe　鶴崎 裕貴† Hiroki Tsurusaki　今野 智明† Tomoaki Konno　河村 圭† Kei Kawamura　内藤 整† Sei Naito

1. はじめに

近年，スポーツ観戦の新たな楽しみ方の一つとして，視聴者が見たいアングルからの映像視聴を可能とする自由視点映像技術が注目を集めている．自由視点映像技術は，あるシーンを複数台のカメラで撮影し，それらの情報を基に実カメラがない視点（仮想視点）を含む任意の視点からの被写体の映像鑑賞を実現する技術である．特に，複数台の実カメラの映像から被写体の 3DCG モデルの形状をボリュームデータとして復元する自由視点方式をフルモデル自由視点[1,2,3]と呼ぶ．従来，このようなフルモデル自由視点の制作処理には膨大な計算処理を要していたが，近年ではリアルタイムで生成が可能な技術[1,2]が提案されてきている．一方，リアルタイム生成の場合には計算コストの少ない処理が求められることから品質を高めることが難しく，品質に関しては向上の余地がある．

フルモデル自由視点映像の品質は，視体積交差法[4]の処理に使われる被写体シルエット画像の抽出精度に大きく影響を受ける．しかしながら，リアルタイム性を志向する場合，多数のカメラ映像に対して、高速にシルエット抽出を行う必要があるため，抽出精度を高めることが困難であった．加えて，既存のカメラキャリブレーション技術ではカメラの位置や向きを完璧に推定することが困難である[5,6]．そのため，被写体の輪郭形状を誤認識なく抽出できるシルエット抽出技術が存在していたとしても，カメラパラメータ推定の誤差に基づいて，視体積交差法の実施時に被写体の 3D モデルの形状に一部欠損が生じる懸念や，逆に被写体ではない領域がモデル化されてしまう懸念があった．加えて，視体積交差法で生成されるボクセルモデルは，通常 1 ㎝や 2 ㎝などの単位ボクセルサイズでモデルの形状が標本化されることから，標本化誤差の観点でも正確な形状を得ることは困難であった． これらの問題を鑑みるに，視体積交差法の前段ではなく，後段での品質改善アプローチが求められる．

そこで本研究では，まず視体積交差法実施前にシルエット画像の被写体抽出領域を膨張させることで 3D モデルの欠損を低減する．このとき，欠損を低減する副作用として，膨張された 3D モデルの輪郭部分に背景のテクスチャがマッピングされることで， 画質の低下が発生する．例えばスポーツ映像を対象にした自由視点映像であれば，選手の 3D モデルの輪郭部分に芝生のテクスチャが写り込むことが考えられる．この問題を解決するために，提案手法では仮想視点からの映像をレンダリングする際にモデルの輪郭を洗練する処理を行う．具体的には，自由視点映像のレンダリング工程において，テクスチャマッピングに使用される実カメラの画素の色情報と，シルエット抽出時に使用される背景モデルの色情報を比較することで，被写体 3D モデルの輪郭部分を透過する．これによりモデル形状を正しく復元することで，高品質な自由視点映像のレンダリングを実現する．

†株式会社 KDDI 総合研究所(KDDI Research, Inc.)

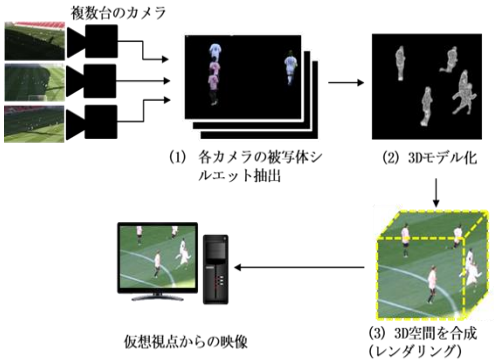


図 1　フルモデル自由視点の制作フロー

2. 提案手法

2.1 自由視点映像視聴システムの概要

本研究における自由視点映像視聴システムの処理の流れを図 1 に示す．処理は以下の 3 工程に大別される．

(1)　シルエット抽出処理 [7]
(2)　3D モデル生成処理 [2]
(3)　レンダリング処理

これらは(1)→(2)→(3)の順にシーケンシャルに処理が成される．基本的には(1)～(3)の各処理に 1 台ずつ別の計算機を割り当てるシステムを想定しており，各処理でリアルタイムが担保できれば，システム全体としてリアルタイム制作を実現することができる．

本研究では(1)のシルエット抽出処理に，高速動作可能な単一ガウス分布ベースの手法[7]を用いた．この手法はリアルタイム自由視点映像制作手法[1]においてもシルエット抽出処理として導入されている．第 1 章で述べた通り，提案手法では [7]の手法で生成された出力シルエット画像に対し，シルエットの被写体抽出領域の膨張（Dilation）処理を実施した．具体的には，シルエット上で被写体が存在すると判定された画素を，周囲$d \times d$画素に拡張することで行われる．

次に，(2)の 3D モデル生成処理には，著者らが過去に提案したリアルタイム自由視点制作手法[2]を用いた．このとき[2]の文献に示される方法で，各カメラの被写体のシルエット画像から被写体のボクセルモデルが計算される．その後，[2]の文献に記載の通り，マーチンキューブ法を用いて被写体ボクセルモデルをポリゴンモデルに変換する．この

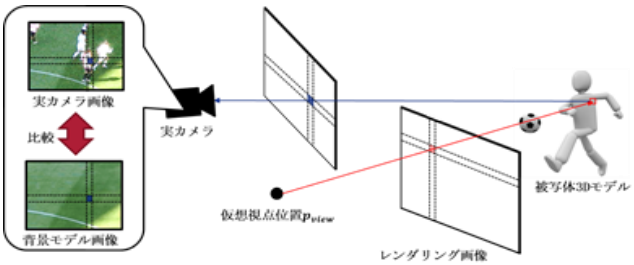


図 2　提案手法のエッセンス

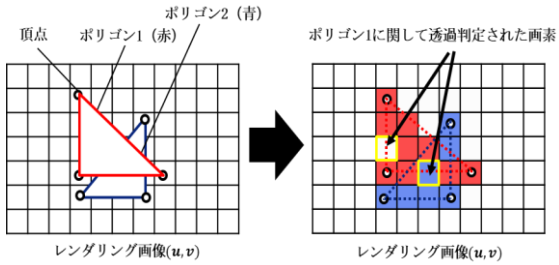


図 3　提案手法における画素の透過判定

とき，同時にポリゴンを構成する各頂点iの遮蔽情報$O_{i,k}$が計算される．$O_{i,k}$は頂点iがカメラkから見て遮蔽されるか否かを表す 2 値の情報である．この遮蔽情報がレンダリング部でのテクスチャマッピング時に使用される．

提案手法は(3)のレンダリング部分を対象にした改善のため，次節でレンダリング部にフォーカスして提案手法の詳細を述べる．

2.2 レンダリング手法

2.2.1 レンダリング手法の概要

前節の手法に基づき被写体 3D ポリゴンモデルを得た後に，仮想視点位置p_{view}から見た 2D 画像上に被写体の描画を行う．この処理をレンダリングと呼ぶ．

提案手法では図 2 のように，レンダリング時のテクスチャマッピングの際に，実カメラの画素を参照する．このときに当該画素上で，実カメラテクスチャと背景モデル画像を比較する．この背景モデル画像はシルエット抽出処理にて実施される背景差分法の過程で得られるガウス分布の平均値を利用する．一般に被写体が存在する場合には比較される画素値には差が生じる．一方，サッカーのフィールド上の芝生などが，3D モデルの生成の不正確さ故に選手のモデルにマッピングされてしまう場合，実カメラテクスチャと背景モデル画像はいずれも芝生が写り込むため画素値の差は小さくなる．このように視体積交差法のためのシルエット抽出だけでなく，レンダリング過程においても実カメラテクスチャと背景モデル画像を比較し，モデル形状を洗練化するのが提案手法の特徴である．

以下に，提案手法のレンダリング手順の概要を述べる．説明においては 3D ポリゴンモデルを構成する各頂点のインデックスをi，ポリゴンのインデックスをnとする．ポリゴンは，どの頂点を結び合わせて構成されるかというポリゴン構成情報を持つ．

【手順 1】ユーザの操作に基づき，レンダリング画像を生成する仮想視点位置p_{view}が決定される．
【手順 2】被写体 3D ポリゴンモデルを 3D 空間上に配置し，仮想視点位置p_{view}から見たレンダリング画像を生成する．そのために，頂点iの 3D ワールド座標位置(x_i, y_i, z_i)に対応する仮想視点位置p_{view}上の 2D 画素位置 (u_i, v_i)を計算する．
【手順 3】仮想視点位置p_{view}から見たレンダリング画像(u,v)上に各ポリゴンP_nの描画を行う．このとき，各ポリゴンP_nを画素(u,v)上に描画する際に透過判定を行い，透過されると判定された画素では透過度を 1 としてポリゴンの描画を行う．方法の詳細は 2.2.2 項で述べる．
【手順 4】手順 3 で施される透過処理を，被写体 3D モデルの輪郭付近のみに限定するために，手順 3 において生成されたレンダリング画像(u,v)上の前景の縁から離れた位置にある画素は透過率を 0 に修正する．この処理の詳細を 2.2.3 項で述べる．

提案手法の主たる貢献は【手順 3】で説明される被写体 3D モデルの透過処理である．前記のカメラパラメータの誤差の問題は，複数台のカメラのシルエットを統合して 3D モデルの形状を得る視体積交差法の実施時に，品質劣化として顕在しやすい．これは，各カメラパラメータの推定誤差がわずかであっても，多数のカメラの結果を統合した際に，蓄積した大きな誤差となってモデル生成の品質に影響を与えるためである．そこで，予めシルエット抽出の際にシルエットの被写体領域を膨張して抽出しておき，レンダリング時に再度背景差分に類似した処理を実施することで，各カメラパラメータの推定誤差が蓄積してモデルが欠損することを防止できる．次項にて，提案手法のポイントである【手順 3】及び【手順 4】の処理を，さらに詳しく説明する．

2.2.2 レンダリング時のポリゴン透過処理

2.2.1 項で示した【手順 3】の詳細を述べる．【手順 3】では，仮想視点位置p_{view}から見たレンダリング画像(u,v)上に各ポリゴンの描画を行う．この処理を図 3 左に示す．各ポリゴンP_nが画素(u,v)上に描画されるとき，各画素にテクスチャマッピングが施される．このとき，最も距離の近いカメラからテクスチャマッピングを施す．提案手法では従来手法[1,2]と同様に，仮想視点位置p_{view}と，実カメラの向き（角度）の差を距離として用いた．ただし，各ポリゴンP_nを構成するいずれかの頂点が遮蔽されている場合，2 番目，3 番目に近い実カメラを参照し，全ての頂点が遮蔽されていない最近傍カメラからテクスチャマッピングを施す．遮蔽情報は 2.1 節で述べた遮蔽情報$O_{i,k}$が利用される．

加えて，提案手法ではこの各ポリゴンP_nを画素(u,v)上に描画する際に透過判定を行い，透過されると判定された画素では透過度を 1（100%）としてポリゴンの描画を行う．ここで画素(u,v)にポリゴンP_nを書き込む際の透過判定の方法について述べる．

まず，判定するレンダリング画像上の画素(u,v)と，描画したい三角形ポリゴン P が存在するとき，P を構成する頂点をV_1, V_2, V_3とする．このV_1, V_2, V_3それぞれの 3D ワールド座標位置を$(x_{v1}, y_{v1}, z_{v1}), (x_{v2}, y_{v2}, z_{v2}), (x_{v3}, y_{v3}, z_{v3})$としたとき，それぞれをテクスチャマッピングに使用する実カメラ上に投影した際の画素位置 $(s_1, t_1), (s_2, t_2), (s_3, t_3)$を計算する．その後，判定する画素$(u,v)$と頂点$V_1, V_2, V_3$の位置

関係に基づき，線形補間により実カメラ上の画素位置(s,t)を得る．その後，以下の判定式を計算する．

$$|I(s,t) - avg(s,t)| < threshold \quad (1)$$

(1)式が正となるとき，三角形ポリゴン P が画素(u,v)に描画される際に，透過率 1(100%)で描画が成される．式(1)の$I(s,t)$はテクスチャマッピングに使う実カメラの画素値，$avg(s,t)$は背景差分法における単一ガウス分布を構成する平均値である．$avg(s,t)$は長時間取得した画像の平均に等しくなるため，人が激しく動くスポーツのシーケンスなどでは人がいない状態の空の舞台の画像となる．$threshold$は手動で決定されるパラメータである．実験では，$I(s,t)$と$avg(s,t)$は共に HSV 色空間で扱い，各色空間で(1)式を計算し，全ての色空間で(1)式を満たすときに当該画素を透過するものとした．また，各ポリゴンの前後関係はデプスバッファに保持され，Z バッファ法に基づいて隠面消去が実施される．

2.2.3　透過領域の決定

2.2.1 項で述べた【手順 4】の詳細について述べる．【手順 3】において生成されたレンダリング画像(u,v)から見て，被写体オブジェクトが存在する画素を前景と見なす．このときレンダリング画像上で，この前景の縁から遠い位置にある画素は透過率を 0 とする．これは，式(1)において$I(s,t)$と$avg(s,t)$の類似性が高い場合に，本来透過されるべきではない領域が透過される懸念があることから，透過を施す領域をオブジェクトの縁付近の領域のみに限定するために実施される処理である．

具体的にはこの処理は，ある画素(u,v)を中心とした$l \times l$[pixels]の矩形領域を考え，矩形領域内の全ての画素にオブジェクトが書き込まれている場合には，縁に遠い画素であると見なして透過率を 0 に戻すことで実施される．

3. 実験

3.1 実験条件

提案手法の有効性を確認するために，サッカーの多視点映像の公開データセット[8,9]を用いて実験を行った．前述のデータセット内で Soccer_Game_Goal として提供されている 10 台のカメラ映像（解像度：1920×1080, 30fps, 10 秒）から自由視点制作を行った．なお公開データセットで提供されるカメラパラメータに誤差が混入しておりモデルに大幅な欠損が生じたことから，これを軽減するために実験ではコート上の白線の交点位置を手動で選択し，事前にサッカーコートの規格から明らかとされる既知の交点の 3 次元位置との対応を基に、手動でカメラキャリブレーションを実施した．また，実験には CPU が Intel Core i7-7700，GPU が NVIDIA RTX 2080Ti，RAM が 64GB の計算機を用いた．またレンダリング画像の解像度は入力画像と同様 1920×1080 とした．

実験では，表 1 に示される 4 種類の条件で自由視点制作を実施した．表 1 の中で条件 D が提案手法に該当する．各条件の詳細を以下に述べる．

条件 A は入力となるカメラ 10 台の被写体シルエットを全て手作業で抽出している．人の目で判断可能なシルエット欠損や誤抽出は全て修正されているため，シルエットの抽出品質は非常に高い．

表 1　各実験条件

条件	シルエット抽出	レンダリング時の透過処理
条件 A	手動	実施せず
条件 B	自動[7](膨張なし)	実施せず
条件 C	自動[7]($d = 5$)	実施せず
条件 D（提案法）	自動[7]($d = 5$)	実施

一方，条件 B～条件 D は[7]の手法に基づき自動でのシルエット抽出が成された．ただし条件 C と条件 D のみシルエット出力時にシルエットの被写体領域に膨張処理を加えている．また，シルエット抽出用の空舞台が本データセットでは提供されていないことから，実行前に 300 フレームのシーケンスに対し，一度[7]の手法でシルエット抽出を実施し，その際に得られる背景モデルを表現するガウス分布の平均値を空舞台画像として利用した．そしてこの空舞台画像を入力に，再度シルエット抽出を実施した．

3D モデルの制作に関しては条件 A～条件 D に関して全て同一の手法[2]を利用した．このときの 3D モデルの単位ボクセルサイズは，文献[2]と同じ 2 ㎝を採用した．

最後にレンダリングに関しては，条件 D のみレンダリング時の 3D モデル透過処理を実施した．条件 D のパラメータは$threshold(h,s,v) = (10,10,10)$，$l$=21とし，経験的に決定された．また，透過処理の際に参照した背景モデルは，前述のガウス分布の平均値として生成される画像を用いた．

なお，スタジアムの背景 3D モデル（ピッチ及び客席）は事前に静的に用意した 3D モデルを配置することで表示を行っている．本技術が生成する対象は選手やボールなどの動的な被写体に限定される．

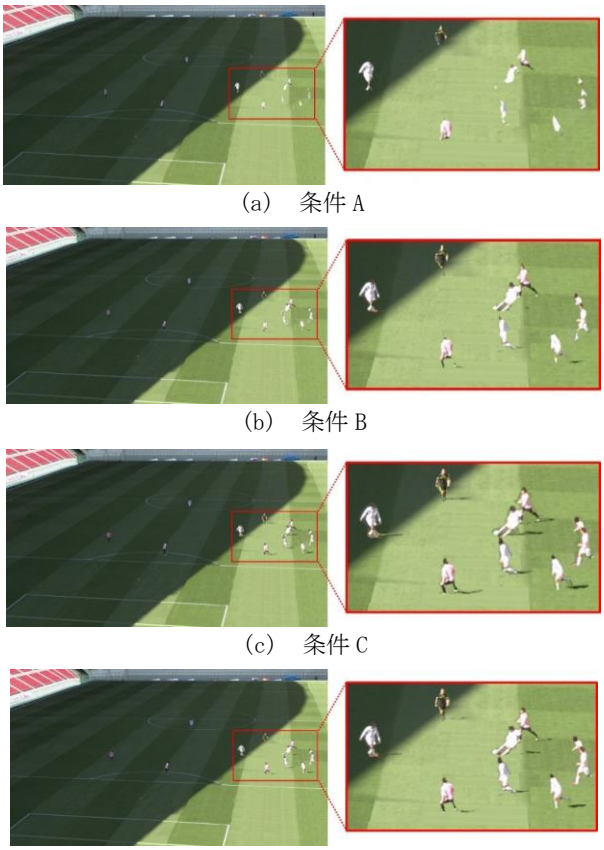
3.2 実験結果

3.2.1 画質の主観評価実験

初めに，提案手法における画質の改善効果に関して検証を行った．結果の一例として，全体 300 フレーム中の 150 フレーム目に関し，特定の視点から見たレンダリング結果を図 4 及び図 5 に示した．

図 4(a)及び図 5(a)の結果より，入力シルエット画像の抽出品質が，手作業によって十分な精度まで高められたとしても，3D モデルの生成品質には限界があることが確かめられた．これはカメラの外部パラメータの推定誤差やレンズ歪の影響が大きく，視体積交差法でモデル生成する際に欠損が発生しているためである．一方，シルエット抽出の時点で既にシルエットに欠損が発生している場合にはモデルが欠けて生成されてしまうことから，視体積交差法前段のシルエット抽出処理では，モデルの欠損を防ぐことに注力することが重要であると考えられる．

次に，図 4(b)及び図 5(b)に示される条件 B では，選手の頭や足の部分の 3D モデルが形成されておらず，欠損が生じている．条件 B の方が，シルエットを手動修正した条件 A よりも欠けが少ないのは，シルエット抽出の判定時に輪郭近傍付近の領域が前景と誤判定されることで，条件 A で使用された前景シルエットよりも若干太めのシルエットが抽出されているためであると考えられる．しかしながら条



(a)　条件 A

(b)　条件 B

(c)　条件 C

(d)　条件 D（提案法）

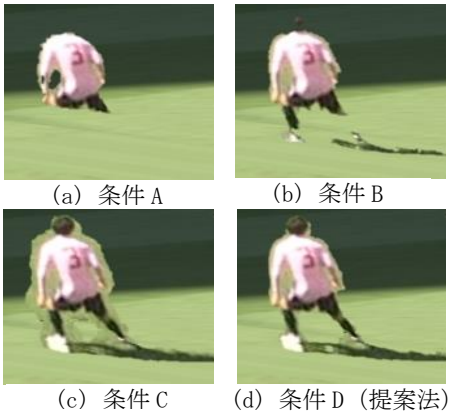
図 4　各条件でのレンダリング結果（引き視点）

件 B においても人物モデルには大きく欠損が生じ，品質が劣化している．

条件 C ではシルエットの被写体が検出された領域の周囲を膨張させているため，欠損を大幅に低減できている．一方，条件 C ではこの膨張の副作用として，実際の被写体のサイズよりも 3D モデルが大きく形成されていることがわかる．特に図 5(c)のように視点位置を被写体に近づけた場合には，選手の周囲に付着する芝生の緑が大きく目立ち，画質を低下させている．一方、図 5(d)は太く抽出されたモデルの輪郭を提案手法の透過処理で洗練しているため，図 5(c)と比較して芝生領域の写り込みを低減できている．また図 5(a)や図 5(b)と比較しても選手のモデルの欠損が少なく，比較的自然な表示が行えている．よって，提案手法を用いて自由視点の視聴品質を向上可能なことが確認された．しかしながら，提案手法においても被写体 3D モデルの輪郭に芝生がマッピングされる問題を完全に防止できていないことから，今後更なる手法の改善を進めると共に，最適なパラメータ等についても検討を行う必要がある．

3.2.2 処理時間の計測

次に，提案手法の処理に基づいて増加した処理時間について検証を行った．レンダリング処理全体の平均フレームレートを 10 秒(300 フレーム)のシーケンスに対して測定した結果，条件 B では 56.72 fps，条件 C で 55.94 fps，条件 D では 55.57 fps となった．なお条件 A は 300 フレームに渡り全てのカメラ画像のシルエットを手動修正することが困難



(a) 条件 A　(b) 条件 B

(c) 条件 C　(d) 条件 D（提案法）

図 5　各条件でのレンダリング結果（寄り視点）

であることから，今回の測定対象からは除外している．このことから提案手法の処理によりわずかに処理時間が増加するものの，元動画のフレームレートである 30fps を大きく上回ることができており，リアルタイムに処理が実施できていることが検証された．また、提案手法ではシルエット抽出の後段にて GPU 実装（CUDA）による被写体シルエットの膨張処理を実施したが，この処理はカメラ 10 台合わせて 1ms 以下であり，生じる処理量増加はわずかである．このことから、本手法はリアルタイム自由視点の品質向上において有効なアプローチとなることが期待される．

4. まとめ

本研究では、レンダリング時の背景モデル参照を用いた自由視点映像の高品質化に関する検討を行った．レンダリングする際の仮想視点の位置に基づいて，背景モデルと画素値が近い場合には 3D モデルに透過を施すことで，リアルタイム自由視点の品質向上に貢献できることを確認した．提案手法が処理時間に与える影響は非常に少なく，本手法を導入しても十分にリアルタイムでレンダリングを実施することが可能であることから，今後リアルタイム自由視点の品質向上に応用されることが期待される．今後，透過判定方法の洗練化や，様々な競技を対象とした実証実験を通し，有効性を検証していく必要がある．

参考文献

[1] K. Nonaka, et al, "Fast Plane-Based Free-viewpoint Synthesis for Real-time Live Streaming," VCIP 2018, pp. 1-4 (2018).
[2] J. Chen, et al, "Fast Free-viewpoint Video Synthesis Algorithm for Sports Scenes", IROS 2019, pp. 3209-3215 (2019).
[3] J. Kilner, et al, "Dual-mode deformable models for free-viewpoint video of sports events, " 3DIM 2007, pp. 177–184 (2007).
[4] A. Laurentini, "The visual hull concept for silhouette-based image understanding," IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 150–162 (1994).
[5] J. Chen, et al, “Sports Camera Calibration via Synthetic Data,” CVPR Workshop 2019, (2019).
[6] L. Wang, et al. "Multi-Camera Calibration with One-Dimensional Object under General Motions," ICCV 2007, pp. 1-7, (2007).
[7] Q. Yao, et al, "Accurate silhouette extraction of multiple moving objects for free viewpoint sports video synthesis," MMSP 2015, pp. 1-6 (2015).
[8] http://www.fujii.nuee.nagoya-u.ac.jp/multiview-data/
[9] R. Suenaga, et al. “Practical Implementation of Free Viewpoint Video System for Soccer Games,” SPIE Electronic Imaging, 3D Image Processing, Measurement and Applications, 9393-15 (2015).