

Visual Hull を応用した自由視点スポーツ映像生成の高効率・高品質化

前田 哲汰 パナヒプル テヘラニ メヒルダド 高橋 桂太 藤井 俊彰

正誤表

訂正箇所	誤	正
タイトル	Visual Hull を応用した自由視点 スポーツ映像生成の高効率・高品質化	Visual Hull を用いた高効率な 高品質自由視点スポーツ映像生成

Visual Hull を応用した自由視点 スポーツ映像生成の効率化・高品質化

前田 哲汰^{1,a)} パナヒプル テヘラニ メヒルダド¹ 高橋 桂太¹ 藤井 俊彰¹

概要:我々は、サッカーなどの広大な空間を用いて行うスポーツを対象として、疎に配置した多数のカメラでシーンを撮影し、visual hull 法を用いてその 3D モデルを生成し、可視化する研究を行っている。Visual hull 法では微小なボクセルの集合によって 3D モデルを表す。扱う空間が広大かつ多数の選手が存在する場合、処理時間の削減やボクセル同士のオクルージョンの効率的な扱いが課題となる。そこで本稿では、処理時間を削減するために、8 分木構造を用いてボクセル群を多重解像度化し、内部を削りながらモデルを生成した。また、オクルージョンを明示的に扱わずに各ボクセルに対して矛盾のない色付けを可能にするために、仮想視点の角度に応じて表示色の変化をフーリエ級数で近似した。この表示色の推定において、周辺のカメラから得た色の中央値を用いることで比較的正しい色付けを行うことができた。

1. はじめに

自由視点映像技術 [1,2] とは、複数の視点から撮影した映像を元に任意の視点からの映像を合成する技術である。この技術はメディアや医療など様々な分野への応用が望まれているが、数ある分野の中でも特にスポーツ中継への応用が期待されている。現在、スポーツ中継では視聴者は放送者の選んだ視点からの視聴しかできないが、応用が実現すれば視聴者が視点を選択しながら視聴することができる。近年、この新たな視聴方式の実現のために、スポーツ中継への自由視点映像生成に対する様々なアプローチが研究されている。サッカーなどのスポーツでは、広大なフィールドを用いる。それゆえ、シーンの 3D モデルの生成には非常に膨大な処理コストが必要となり、実用化は難しい。

そこで、我々は少ない処理コストで自由視点映像生成を行う手法について研究を行っている。この研究ではスポーツ中継の放送と並行した処理によって高品質な映像を生成することを目的としている。以前、我々は visual hull [3-5] に 8 分木構造を導入した 3D モデル生成手法 [6] やフーリエ級数展開を用いた色付け方式 [7] について報告したが、モデル生成効率や色付けの品質に課題があった。

本稿では、これらの課題を解決するために 2 つの手法を報告する。1 つ目はモデルの生成効率を向上する手法である。[6] によるモデル生成ではボクセル探索に無駄があっ

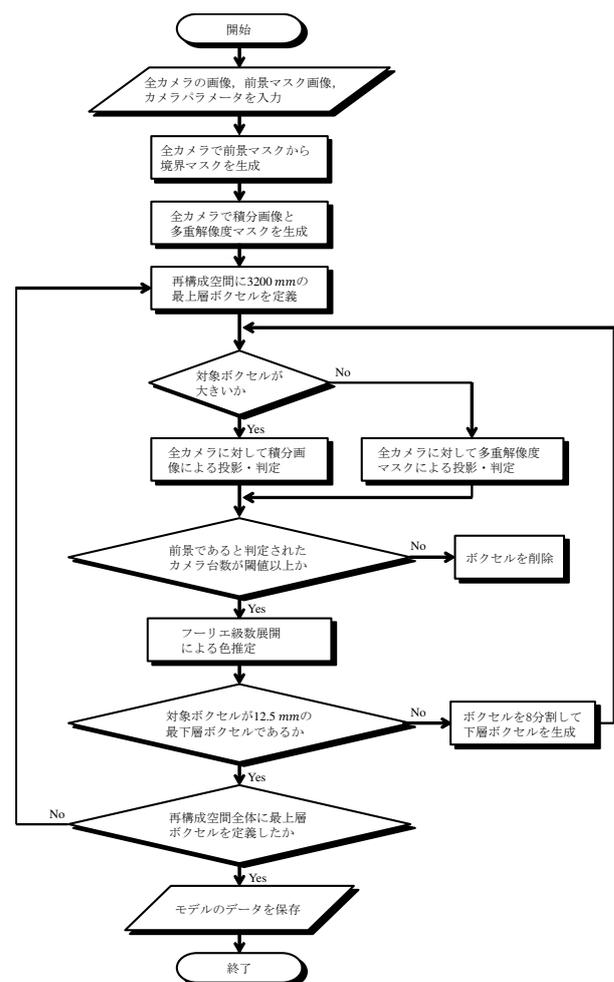


図 1 モデル生成のフローチャート

¹ 名古屋大学大学院 工学研究科
Furo-Cho, Chikusa-Ku, Nagoya, Aichi 464-8603, Japan
^{a)} maeda@fujii.nuee.nagoya-u.ac.jp

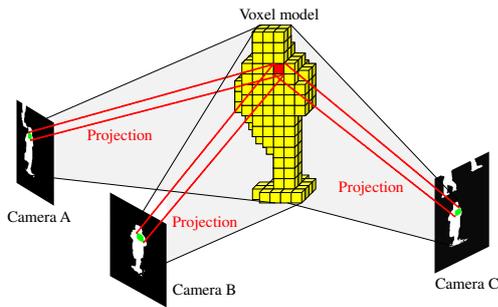


図2 Space carving method によるボクセルモデル

たため、さらに効率の良い探索手法について述べる。2つ目はモデルへの色の割り当ての品質を向上させる手法である。[7]ではモデルの色の推定の精度が悪かったため、精度のよい推定法について述べる。

2. 自由視点映像生成手法

本節では、以前我々が報告している自由視点映像の生成手法の概要を説明する。我々は仮想視点の移動範囲に制限がないレンダリングに着目し、モデルベースドレンダリングの手法を採用している。しかし、モデリングの際の表面形状や反射特性の推定には多大なコストが必要となるため、本手法の3Dモデルは形状推定に留めた。図1に3Dモデルの生成のフローチャートを示す。以下の小節ではモデル生成のための各処理について説明する。

2.1 多重解像度 Visual hull

我々は3Dモデルの生成に visual hull 法（視体積交差法）を用いた。Visual hull 生成は space carving method [8]によって行う。まず、各カメラで得た画像を元に前景マスク画像を用意する。続いて、任意の大きさのボクセルが隙間なく敷き詰められているボクセル空間を定義する。ボクセル空間中の全てのボクセルに対して以下の処理を行う。ボクセルを三次元座標から各カメラの二次元座標へとカメラパラメータを用いて投影する。投影の式は以下に示す。

$$\mathbf{x}' \sim P\mathbf{X}' \quad (1)$$

ここで、 \mathbf{x}' 、 \mathbf{X}' は投影先の2次元座標と投影前の3次元座標を意味し、それぞれ同次座標表現である。また、 P はカメラパラメータからの射影行列である。このようにして投影されたボクセルが各カメラにおける前景マスクの前景領域と重なっているかを判定する。これらの処理を本稿では投影・判定処理と呼ぶ。重なっていると判定されたカメラ台数が事前に設定した閾値以上であればそのボクセルを残し、そうでなければ削除する。この処理を本稿では閾値処理と呼ぶ。このようにして最終的に残ったボクセル群が選手のモデルを形作る。この様子を図2に示す。

Space carving method において、鮮明なモデルを生成するためには空間を埋め尽くすボクセルのサイズを小さくす

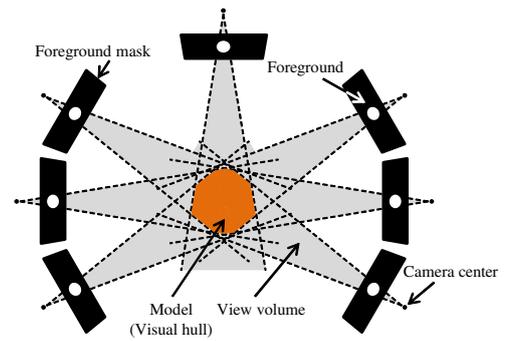


図3 Visual hull 生成

ることで解像度を上げる必要がある。しかし、ボクセルを小さくすると相対的に空間内のボクセル数は増えるため処理の回数も増える。そこで、処理回数を減らすために8分木構造を導入する。8分木構造では初めに空間に定義するボクセルのサイズを大きくし、それぞれに対して投影・判定処理を行う。その後、閾値処理によって残ったボクセルは8つのボクセルに分割する。それぞれの分割されたボクセルに対して同様の処理を繰り返すことによって、選手のいない領域は大きなボクセルの段階で削除し、選手のいる領域は細かく探索する。

8分木構造を導入することで空間内を効率よく探索しながらボクセルモデルを生成できる。しかし、大きなボクセルの投影・判定処理では、微小なボクセルの処理と比べると処理コストが大きくなる。そこで、我々は積分画像と多重解像度マスクを用いた投影・判定処理の近似手法を提案した。この手法の詳細については[6]を参照してほしい。

上記のボクセルモデル生成を多重解像度 visual hull と呼ぶ。ボクセルを多重解像度化することによって効率良くモデル生成を行うことができる。本稿では、親ノードを一辺が3200 mmのボクセルとして最上層の第0層とする。分割される毎に一辺の長さは1600 mm、800 mmと半分になり、それぞれを第1層、第2層のボクセルとして扱う。また、最下層は第8層とし、一辺の長さは12.5 mmとする。

Visual hull 法では各カメラでの物体領域に対する視体積の積をとるため、生成されるモデルは図3のような中身の詰まったボリュームモデルとなる。そのため、描画に影響のないモデルの内部に対して細かく探索を行ってしまう。また、モデルを形成するボクセルとして残ってしまうため、後述する色の割り当てや最終的な描画に関しても無駄な処理を行わなければならない。そこで、3.1節ではこの問題を解決するための手法を提案する。

2.2 フーリエ級数展開による色付け方式

我々は[7]で、求めたボクセル群に色を割り当てる方式を提案した。本節ではその方式について詳しく述べる。

本稿ではシーンの3Dモデルを反射特性を持たないボクセルモデルとしている。しかし、視点の角度によってモデ

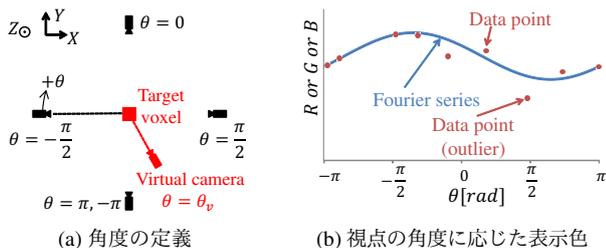


図4 水平方向の角度に応じた表示色の定義

ルの色は変化するの自然であるため、一つのボクセルの色を一意に決めるべきではない。したがって、仮想視点の角度に応じて各ボクセルが表示する色を変化させる必要がある。そこで本稿では、少ない係数から成るフーリエ級数を用いることで各ボクセルの表示色を表現する。フーリエ級数は周期 2π の周期関数であるため、これによって水平方向に関して全方向へ異なる色を割り当てることができる。

まず、各ボクセルに対して以下の式(2)を解きフーリエ係数 a_0 , a_1 , b_1 を求める。

$$\arg \min_{\mathbf{F}} \|\mathbf{C} - \mathbf{A}\mathbf{F}\|_2^2 \quad (2)$$

$$\mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \frac{1}{2} & \cos \theta_1 & \sin \theta_1 \\ \frac{1}{2} & \cos \theta_2 & \sin \theta_2 \\ \vdots & \vdots & \vdots \\ \frac{1}{2} & \cos \theta_n & \sin \theta_n \end{bmatrix}, \mathbf{F} = \begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix} \quad (3)$$

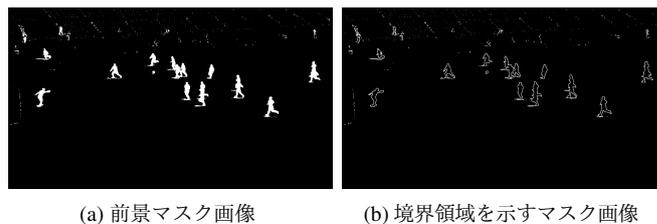
ここで、 c_i は対象のボクセルに対する各カメラからの色(R_i, G_i, B_i)、 θ_i は絶対座標と対象のボクセルとカメラが成す水平方向の角度である。また、 n はデータの取得回数である。さらに、 $\mathbf{a}_0 = (a_{0R}, a_{0G}, a_{0B})$, $\mathbf{a}_1 = (a_{1R}, a_{1G}, a_{1B})$, $\mathbf{b}_1 = (b_{1R}, b_{1G}, b_{1B})$ である。次に以下の式(4)から仮想視点に対して表示する色を決定する。

$$\mathbf{f}(\theta_v) = \frac{1}{2}\mathbf{a}_0 + \mathbf{a}_1 \cos \theta_v + \mathbf{b}_1 \sin \theta_v \quad (4)$$

ここで、 $\mathbf{f} = (f_R, f_G, f_B)$ は仮想視点に対して対象のボクセルが表示する色、 θ_v は絶対座標とボクセルと仮想視点と成す水平方向の角度である。このとき、水平方向の角度については図4(a)に示すように、上方から見て時計回りになるように定義する。

このような色付け方式を用いると、角度を変数とした式で表示色を表現することができる。モデルの色の変化は緩やかであると仮定し、フーリエ係数はそれぞれの色に対して a_0 , a_1 , b_1 のみとした。このときのフーリエ級数による仮想視点の角度に応じた表示色の例を図4(b)に示す。

しかし、上記の色付け推定では入力画像の一部で選手同士が重なってしまった場合、それらがオクルージョンとし



(a) 前景マスク画像 (b) 境界領域を示すマスク画像

図5 前景マスク画像とその境界領域を示すマスク画像

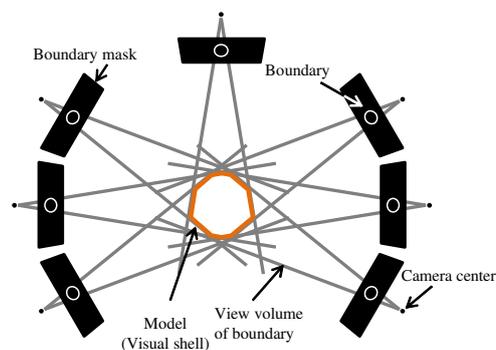


図6 境界領域の視体積により生成される visual shell

で影響を及ぼし、生成されたモデルに間違っただけの色がついてしまう。また、推定精度が低いためモデル全体のコントラストも低下してしまう。そこで、3.2節ではオクルージョンの影響を抑えた精度の良い色付け推定について提案する。

3. 提案手法

3.1 Visual shell によるモデル内部の削除

本節では visual hull 法よりも効率の良いボクセルモデル生成手法について提案する。これは、モデルの内部を削除しながら探索することで高速化を図る手法である。また、この手法を用いるとモデルを形成するボクセルが減るため、後述する色の割り当てにおいて処理量を減らし、最終的なモデルのデータ量を削減することができる。

まず各前景マスクにおいて、背景領域に近い前景領域の画素を選択し、図5(b)に示すような境界領域を示すマスク画像を生成する。このとき、選択する画素は周辺 σ 画素に背景画素が存在する前景画素であり、本稿では $\sigma = 6$ とした。次に、各カメラでの境界領域における視体積の和を取り、それと visual hull との積を取る。これによりモデル内部のボクセルは削除され、描画に影響のあるモデル表面に位置するボクセルのみが残る。これは図6に示すようなサーフェスモデルである。これを visual shell と名付ける。

このとき、visual hull 法によるモデル生成を終えた後に内部の削除を行うと、モデル生成の計算コストは単純に増大する。しかし、8分木探索によって階層的に削除を行うと、疎なボクセルにおいて内部が削除されるため処理するボクセル量が減る。つまり、生成と色の割り当てにおいて計算コストを削減し、モデルの容量を抑えることで描画に

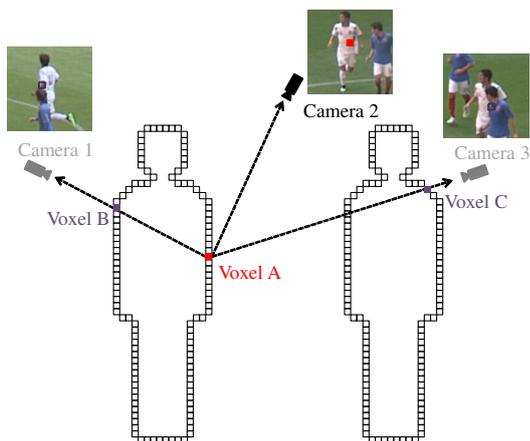


図7 オクルージョンの影響を受けるカメラ

においても処理量を減らすことができる。

3.2 ローカル中央値を用いた色推定

各カメラにおいてボクセルの色の取得は、ボクセルの中心座標をそのカメラの画像に投影し、その投影された座標の画素を参照することで行う。しかし、各ボクセルにおいて正しい色を取得できないカメラが存在する。そのようなカメラは以下の二種類のオクルージョンの影響を受けている。一つはその選手自身の他のボクセルによるオクルージョンである。例えば、図7に示したカメラ1がボクセルAの色を参照しようとした場合、ボクセルAを含むモデル表面はカメラ1の反対側に位置するため手前にあるボクセルBの色を取ってしまう。もう一つは、他の選手によるオクルージョンである。例えば、図7に示したカメラ3がボクセルAの色を参照しようとした場合、他の選手がボクセルAとカメラの間に入るため手前の選手のボクセルCの色を取ってしまう。つまり、図7においてボクセルAの色を正しく取得できるのはカメラ2のみとなる。2.2節で提案した色付け方式を用いると、前者の影響は小さく抑えることができるが後者の影響は残ってしまう。そこで本小節では、二種類のオクルージョンの影響を最小限に抑える色推定手法を提案する。

オクルージョンの影響を抑えて正しい色を推定するために重み付き最小二乗を用いてロバスト推定を行う。有名な重み付きのロバスト推定として *biweight* 推定法がある。*Biweight* 推定法では最小二乗法で求めた関数を元に重みを付ける。関数から値の離れたデータ点の重みは0とすることで次の推定の際には除く。さらに残ったデータ点と関数の差が小さければ大きな重みを付け、差が大きければ小さな重みを付ける。重みを付けたデータ点を再度最小二乗法で解くことによってオクルージョンによる外れ値の影響を抑えることができ、重み付けと推定を繰り返すことで正しい関数を得ることができる。

しかし、目的関数が複数の極値を持つ場合、反復を用いる

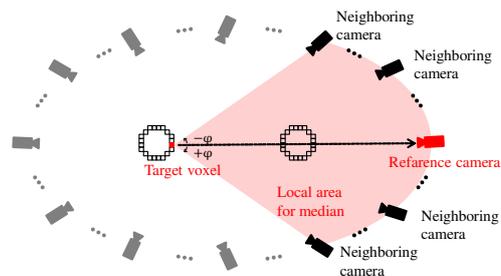


図8 重みを付ける参照カメラと周辺カメラ

手法では最適ではない局所解に収束する恐れがある。つまり、反復を用いるとオクルージョンの影響や投影のずれによって間違った関数を求めてしまうことがある。したがって、そのような問題を解決する重み付けについて述べる。

以下の処理を全てのボクセルに対して行う。まず、各カメラをそれぞれ参照し、参照カメラの周辺のカメラで得た色について中央値を求める。このとき、周辺カメラの定義は参照カメラから水平角度 $\pm\phi$ 以内に存在するカメラとし、フィールドを上から見た図8にこの様子を示す。ここで、本稿では $\phi = \frac{\pi}{3}$ とした。次に、各データ点の重みを以下の式(5)で決定する。本稿では色の推定は基本的にR, G, Bそれぞれ独立で行うが、重みについては共通とする。

$$w_i = \begin{cases} \left\{1 - \left(\frac{d_i}{J}\right)^2\right\}^2 & (d_i < J) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

$$d_i = \|c_i - m_i\|_2^2 \quad (6)$$

ここで、 J は誤差の許容範囲のパラメータ、 m_i は参照カメラ周辺のローカル中央値(m_{iR}, m_{iG}, m_{iB})である。また、 i はデータのインデックスである。本稿では $J = 40$ とした。続いて、次式(7)によってフーリエ係数を求める。

$$\arg \min_F \|W(C - AF)\|_2^2 \quad (7)$$

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \quad (8)$$

以上による推定では、中央値を用いた重み付けにより外れ値の影響が小さい。また、*biweight* 推定法と比較すると、反復を行わないため間違った値を取ることは少ない。

4. 実験

本章では、提案手法の有効性を確かめるために、実際に自由視点映像生成の実験を行った。入力には図9に示す14視点から撮影された映像とそれぞれの前景マスク画像、カメラパラメータを用いた。映像の解像度は4k(4096×2160)である。実験に使用したフレームは1フレームとした。

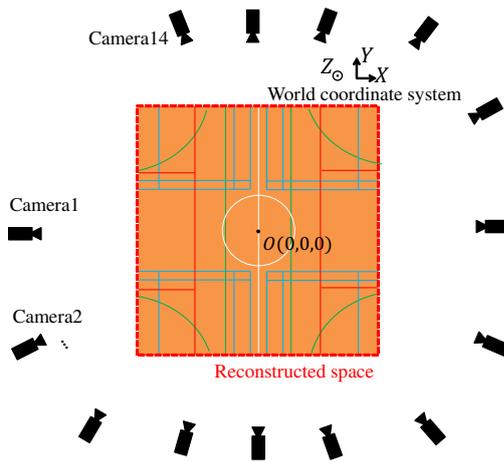


図9 再構成空間と世界座標

実験を行った PC の構成を表 1 に示す。自由視点映像の描画にはフリーのグラフィックライブラリ OpenGL を用いた。本稿では、CUDA 等の GPU プログラムは使用しない。

本節ではモデルの生成コストと映像の品質についてそれぞれ評価するが、ここで比較手法について説明する。比較手法のボクセルの探索は 8 分木で行う。投影・判定処理と閾値処理は提案手法と同様に行う。さらに、ボクセルが表示する色は仮想視点の両側に位置するカメラで得た色をブレンドしたものとした。つまり、モデルの生成時には各カメラの水平方向の角度と取得した色をモデルの情報として保存し、描画の際に仮想視点の角度を元に色を計算した。ブレンドの比率は以下の式を用いた。

$$g(\theta_v) = \frac{\beta}{\alpha + \beta} c_A + \frac{\alpha}{\alpha + \beta} c_B \quad (9)$$

ここで、 $g = (g_R, g_G, g_B)$ はブレンドされた色を表す。また、 $\alpha = \theta_A - \theta_v$ 、 $\beta = \theta_v - \theta_B$ とし、 A 、 B はそれぞれ仮想視点の左右のカメラを意味する。

4.1 モデルの生成コスト

まず、8 分木と 3.1 節を用いたモデル生成にあたり、ボクセル探索回数がどの程度削減されるのか実験を行った。12.5 mm のボクセルの全探索、3200 mm のボクセルを最上層とし 12.5 mm のボクセルを最下層とした 8 分木探索、visual shell を導入した 8 分木探索についてそれぞれ比較した。全探索は 2.68×10^8 回であったのに対して、比較手法は 7.34×10^5 回、提案手法では 5.21×10^5 回であった。

次に、色の割り当てについて、一つのボクセルに対する処理時間に関して実験を行った。このとき、比較として biweight 推定法を用いて色推定を行った際の計算時間測定した。biweight 推定法では推定の繰り返し回数を 1~4 回で変化させた際の処理時間をそれぞれ示す。提案手法は 8.57×10^{-2} ms となり、Biweight 推定法ではそれぞれ 1.02×10^{-1} ms, 1.27×10^{-1} ms, 1.55×10^{-1} ms, 1.83×10^{-1} ms となった。Biweight の繰り返し回数が 1 回のとき提案

表 1 PC の構成

OS	Windows 7 Professional
CPU	Intel Xeon CPU E5-2690 v3 2.60GHz
Main Memory	64GB
GPU	NVIDIA TITAN X

手法との差は小さいが、繰り返し回数を増やすと提案手法の方が大幅に短い。

続いて、比較手法と提案手法において、モデル生成にかかる合計時間を比較した。比較手法は 7.45×10^4 ms だったのに対し、提案手法は 3.59×10^4 ms だった。比較手法は 8 分木による探索を行っているが、モデルの内部のボクセルに対して細かく探索してしまうため、生成コストが大きくなった。一方で、提案手法では色の割り当てにコストを割いてはいるが、無駄な探索、色付けをしないため生成コストは小さい。

また、残された最下層ボクセルの数は比較手法が 5.22×10^5 点であったのに対して提案手法は 2.53×10^5 点であった。提案手法では visual shell によってモデルのボクセル数を削減し、モデルのデータ容量を減らすことができるということがわかった。

4.2 映像の品質

比較手法と提案手法によって生成されたモデルを用いた自由視点映像に対して評価した。

他の選手のオクルージョンの影響がある視点を仮想視点とした映像の結果を図 10 に示す。また、平均色の結果と色の推定に biweight 推定法を用いて推定を 4 回繰り返した結果を同様に示した。このとき、各データ点の重みは式 (5) により求めるが、繰り返し 1 回目の推定では $J = 40$ とし、回数が増すごとに J を 5 ずつ減らすことで徐々に推定精度を上げた。このとき、各データ点の重みは式 (5) により求めるが、 d_i については以下の式 (10) を用いて決定した。

$$d_i = \|c_i - f_i(\theta_i)\|_2^2 \quad (10)$$

ここで、 f_i は以前に求めた係数を用いた式 (4) に θ_i を代入して求め、新たな係数については式 (7) によって求めた。

平均色を用いると、多くのカメラがオクルージョンの影響を受けているためコントラストが大幅に下がる。比較手法では、手前の人物の色が奥の人物に割り当てられてしまうため非常に不自然な映像となる。一方で、提案手法はそのようなオクルージョンの影響を受けないような色付けが可能となっている。また、biweight 推定法でもオクルージョンの影響を消すことはできたが、誤った色に収束してしまった箇所が見られた。

誤った色に収束したボクセルを調べた。図 10(d) の赤い正方形内に存在する 1 つのボクセルが角度に応じて表示する色を図 11 に示す。上からカメラ番号、各カメラで得たボクセルの色、ローカル中央値、各カメラの平均による色付

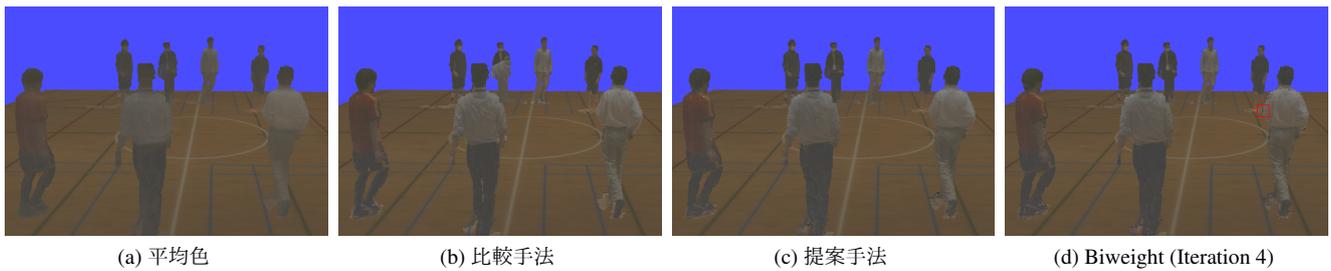


図 10 生成される自由視点映像

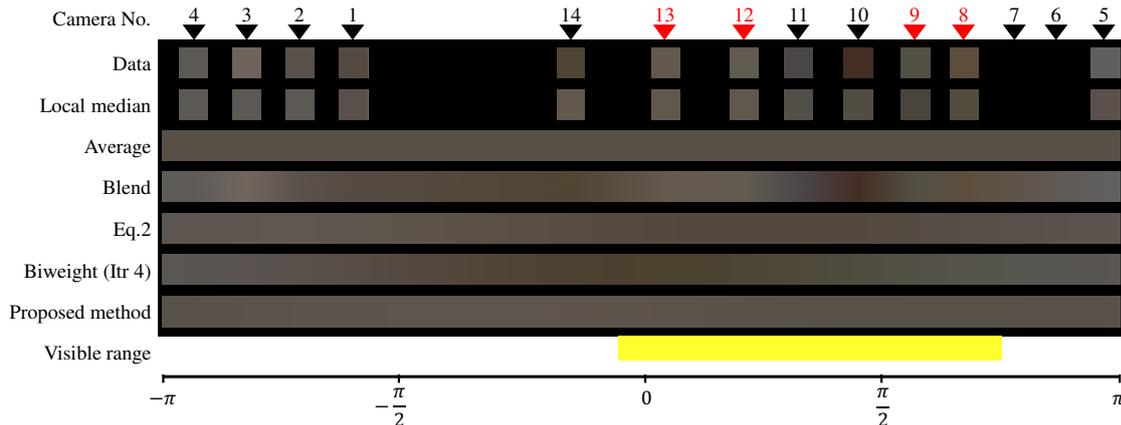


図 11 一つのボクセルが仮想視点の角度に応じて表示する色

け、ブレンドによる色付け、式 (2) による色付け、biweight (繰り返し 4 回) による色付け、提案手法による色付け、そのボクセルを確認できる仮想視点の角度の範囲である。カメラ番号が赤いカメラで得た色は正解の色であり、カメラ番号が黒いカメラで得た色はオクルージョンの影響を受けた色である。ブレンドは各カメラで得た色を線形に補間しているためカメラ 10, 11 のオクルージョンの影響を大きく受けている。また、式 2 による色付けではそれらの影響が残ってしまう。Biweight (繰り返し 4 回) ではカメラ 10, 11 のオクルージョンの影響は少ないが、カメラ 14 のオクルージョンの影響が大きいため $\theta = 0$ 付近の色が黒ずんでいる。一方、提案手法ではオクルージョンの影響が小さく、正解の色に近い色付けであることがわかる。

5. まとめ

Visual hull 法を用いたボクセルモデル生成において、内部を削除しながらモデルを生成する visual shell を提案し、探索数の削減や無駄な色付け処理の防止をすることで処理コストを抑えることができた。さらに、生成されるモデルのデータ容量も削減することができた。

また、フーリエ級数を利用したボクセルモデルへの色付け方式に対して、ローカル中央値から重み付けを行う推定法を提案し、biweight 推定法による推定と比較して正しい色付けをすることができた。

今後の取り組みとしては、フーリエ級数以外の関数を用いた色付け方式や、オクルージョンの影響をさらに抑えることができる推定法が挙げられる。

参考文献

- [1] T. Fujii, M. Tanimoto, "Free-viewpoint TV system based on ray-space representation," Proc. of the SPIE ITcom, Vol. 4864, pp. 175–189, 2002.
- [2] M. Tanimoto, M. Panahpour Tehrani, T. Fujii, T. Yendo, "FTV for 3D spatial communications" Proc. of IEEE, 100(4), pp. 905–917, 2012.
- [3] B. Baumgart, "Geometric Modeling for Computer Vision," PhD thesis, Stanford University, 1974.
- [4] G. K. M. Cheung, S. Baker, T. Kanade, "Visual Hull Alignment and Refinement Across Time: A 3D Reconstruction Algorithm Combining Shape-From-Silhouette with Stereo," Proc. of CVPR, 2, II-375–82, 2003.
- [5] 中村, 斎藤, "Visual Hull を利用した多視点画像からの仮想視点画像生成," 電子情報通信学会 2002 総合大会, D-12-134, 2002.
- [6] T. Maeda, R. Suenaga, K. Suzuki, M. P. Tehrani, K. Takahashi, T. Fujii, "Free Viewpoint Video for Sports Events Using Multi-resolution Visual Hull and Micro-facet Billboard-ing," Proc. of SISA 2016, SS3-5, pp. 191–196, Sep. 2016.
- [7] 前田, テヘラニ, 高橋, 藤井, "自由視点 Visual hull のためのフーリエ級数展開を用いた仮想視点の角度に応じたボクセル表示色の設定," Proc. of IMPS 2016, P-5-09, pp. 198–199, Nov. 2016.
- [8] K. N. Kutulakos, S. M. Seitz, "A Theory of Shape by Space Carving," IJCV, Vol. 38, Issue 3, pp. 199–218, 2000.