

テクスチャ合成を利用した多視点画像からの 高品質な自由視点画像生成法

沢田奈保子[†] 高橋実^{††} 藤本忠博[†]

本研究では、対象物体を周囲から撮影した複数の参照画像をもとに、テクスチャ合成を利用して、任意の視点位置から見た高品質な自由視点画像を生成する手法を提案する。本手法では、まず、複数の参照画像から視体積交差法とボクセルカラーリング法を用いて物体のボクセルモデルを生成し、このモデルを参照画像とは異なる視点位置からレンダリングすることで仮の自由視点画像を生成する。そして、この画像と元の参照画像を用いたユーザ制御型のテクスチャ合成により、仮の自由視点画像の品質を向上させる。

High-quality Free-viewpoint Image Generation from Multi-view Images Using Texture Synthesis

Naoko Sawada[†] Minoru Takahashi^{††} Tadahiro Fujimoto[†]

In this paper, we propose a method to generate high-quality free-viewpoint images from reference images on which a target object is captured from surrounding camera positions. This method utilizes a texture synthesis technique. First, the voxel model of the object is constructed from the reference images by the volume intersection method and the voxel coloring method. Then, the voxel model is rendered for an arbitrary free-viewpoint to generate a temporary free-viewpoint image. Finally, the quality of the temporary image is improved to generate a final free-viewpoint image by the user-controlled texture synthesis technique with a reference image.

1. はじめに

対象物体を複数の異なるカメラ位置から撮影した参照画像（多視点画像）を用いて、任意の自由視点位置から見た自由視点画像を生成するイメージベース CG 技術が盛んに研究されている。その代表的なアプローチの一つとして、参照画像から物体の3次元形状をボクセルモデルとして復元し、自由視点からレンダリングする方法がある[1]。しかし、この方法では、通常、ボクセル空間の解像度が制限されることから、3次元形状の復元精度が低く、結果として自由視点画像の品質が低くなってしまいう問題がある。また、一般に、参照画像の枚数が少ない場合には、参照画像を撮影したカメラ位置から自由視点位置が離れるほど、自由視点画像の品質は低下する。そこで、それらの問題を解決するため、本研究では、ボクセルモデルとして復元した3次元形状を自由視点に対してレンダリングした画像を“仮の”自由視点画像とし、その画像に対してテクスチャ合成を行なうことで、品質を改善した自由視点画像を得る手法を提案する。本手法は、ボクセルベースの視体積交差法[1]とユーザ制御型テクスチャ合成法[2]を用いた先行研究[3]に基づき、ボクセルモデルによる3次元形状の復元に視体積交差法とボクセルカラーリング法[4]を用い、テクスチャ合成には PatchMatch 法による近似ニアレストネイバー探索[5]とユーザ制御型テクスチャ合成法を用いる。これにより、高品質な自由視点画像を効率的に得ることができる。

2. 本手法の概要

本手法の処理の流れを図1に示す。まず、対象物体を異なるカメラ位置から撮影し、複数枚の参照画像を得る。次に、参照画像に対して視体積交差法とボクセルカラーリング法を適用することで、ボクセルモデルで3次元形状を復元する。そして、ボクセルモデルを自由視点からレンダリングして仮の自由視点画像を得る。最後に、仮の自由視点画像に対して PatchMatch 法とユーザ制御型テクスチャ合成法によるテクスチャ合成を行なうことで、高品質な自由視点画像を生成する。以降の節で各処理について詳しく説明する。

[†] 岩手大学
Iwate University
^{††} NTT コムウェア
NTT Comware

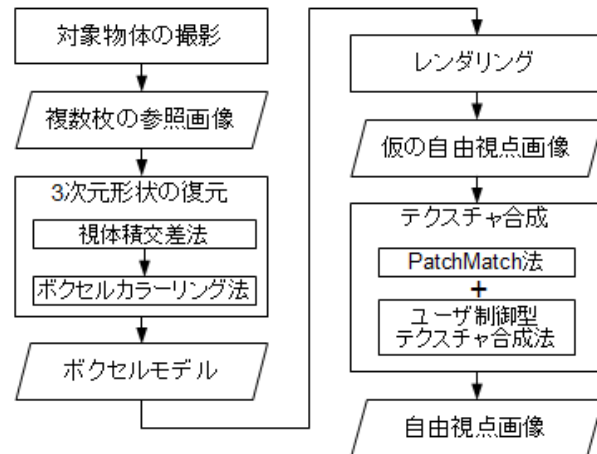


図 1 本手法の処理の流れ

3. ボクセルモデルの生成

3.1 3次元形状の復元と着色

複数枚の参照画像から、視体積交差法とボクセルカラーリング法を用いることで、対象物体の3次元形状を復元する。

はじめに、視体積交差法により、ビジュアルハルと呼ばれる3次元形状を復元する(図2)。ビジュアルハルとは、参照画像上の物体のシルエットを用いて定義される3次元形状であり、実際の物体形状を内包する近似形状となる。これは、各カメラ位置から物体シルエットを3次元空間上に投影して得られる視錐体について、すべてのカメラの視錐体の共通領域として定義される。視体積交差法では、まず、参照画像の各画素を物体と背景のいずれが写っているかにより物体画素と背景画素として2値化する。次に、物体を内包するボクセル空間を設定し、個々のボクセルを各参照画像上に投影する。そして、すべての参照画像の物体画素上に投影されるボクセルはビジュアルハルの内部のボクセルとして残し、一方、背景画素に投影される参照画像が一つでもあるボクセルは外部のボクセルとして除去する(図3)。

ビジュアルハルによる近似形状は参照画像上の物体シルエットのみに基づくため、

復元精度が低く、特に、参照画像上に現れない凹形状を復元できないという問題がある。そこで、参照画像の色情報を用いたボクセルカラーリング法により、ビジュアルハルのボクセルモデルの復元精度の改善を行う(図4)。ボクセルカラーリング法により改善された3次元形状はフォトハルと呼ばれる。まず、上記の処理でモデル中に残されたボクセルのうち、表面に位置する各ボクセルについて、どのカメラから可視であるかを判定する。そして、可視である各カメラの参照画像上に表面ボクセルを投影し、投影された位置の画素の色を得る。すべての可視カメラで得られた色が一致していれば、その表面ボクセルを残し、一致していなければ除去する。その後、新たに表面上に現れたボクセルについても同様の処理を行い、この処理を除去されるボクセルがなくなるまで繰り返す。最終的に残ったボクセル群が物体のフォトハルを表すボクセルモデルとなる。このモデルの各表面ボクセルには、上記の処理で各可視カメラの参照画像から得られた色の平均色を与える。

上記の可視カメラの判定については、各ボクセルにxyz座標に対応させた固有の色ID(RGB値)を与え、RGB値によって各ボクセルを特定できるようにする方法を用いる。まず、カメラごとに、そのカメラ位置から現在のボクセルモデルを色IDでレンダリングして判定用画像を生成しておく。次に、そのボクセルモデル上の表面ボクセルVごとに、各カメラの判定用画像上に投影を行う。そして、判定用画像上の投影された位置の画素の色がボクセルVの色IDと一致した場合は、そのカメラからボクセルVが可視であると判定し、一致しない場合は不可視と判定する。なお、各ボクセルに固有の色IDを与える都合上、ボクセル空間の最大解像度は 256^3 となる。

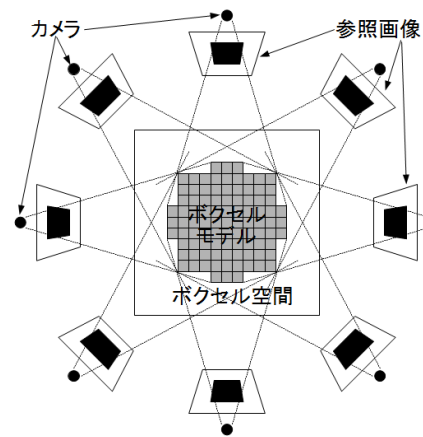


図 2 視体積交差法(全体図)

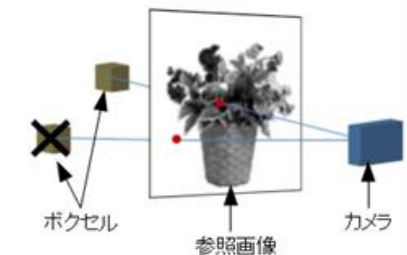


図 3 視体積交差法(ボクセルの除去)

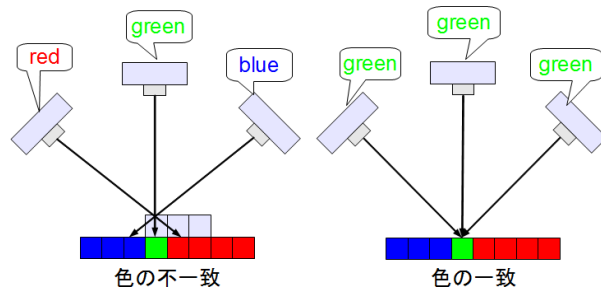


図 4 ボクセルカラーリング法

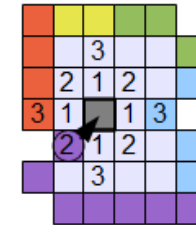


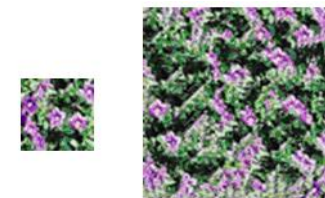
図 5 色の補足

3.2 色の補足

3.1 節の方法によるボクセルモデルの生成において、特に、対象物体の形状が複雑な場合や、参照画像を撮影するカメラ位置が制限されている場合には、いずれのカメラからも可視でない表面ボクセル（不可視ボクセル）が生じる。このとき、いずれかのカメラから可視である表面ボクセル（可視ボクセル）に対しては色が与えられるが、一方、不可視ボクセルには色を与えることができない。そのため、不可視ボクセルに対しては、そのボクセルから最も近い可視ボクセルを探し、その可視ボクセルの色を与えることにした。図 5 は、この方法を 2 次元で示したものである。各四角形がボクセルを表し、濃い灰色と薄い灰色が不可視ボクセル、その他の色が可視ボクセルを表す。中央の濃い灰色の不可視ボクセル（注目ボクセル）に対して、最も近い左下の可視ボクセルの色が与えられる。図中のボクセル上の数字は注目ボクセルからの距離順を表し、この順に最も近い可視ボクセルを探索する。実装上は、この距離順に注目ボクセルから周辺ボクセルへのボクセル座標の差分（相対座標）を事前にテーブル化しておくことで、探索処理を高速化している。

4. テクスチャ合成

与えられた入力テクスチャ（入力画像）に基づき、そのテクスチャと同じパターンを持つ任意サイズの出力テクスチャ（出力画像）を生成する処理をテクスチャ合成という（図 6）。本手法では、PatchMatch 法による近似ニアレストネイバー探索を利用したユーザ制御型テクスチャ合成法を用いる。



入力テクスチャ 出力テクスチャ

図 6 テクスチャ合成

4.1 PatchMatch 法

本手法で用いるテキストチャ合成では、2つの画像（後述のように、実際には、入力画像、出力画像、目的画像の3つの画像を用いるが、処理上、出力画像と目的画像を1つの画像として扱う）について、一方の画像の画素ごとに、その画素を中心とした正方形の画素群（パッチ）に最も類似するパターンを持つパッチを他方の画像内で探索する必要がある。この最類似パッチはニアレストネイバー（Nearest Neighbor: NN）と呼ばれる。本手法では、この探索法として、近似的なニアレストネイバーを高速に探索する PatchMatch 法[5]を用いる。画像 A 内の座標 \mathbf{a} の画素 P_a を中心としたパッチ S_a と類似するパッチを画像 B 内で探索して得られたパッチ S_b の中心画素 P_b の座標を \mathbf{b} とする。このとき、オフセット $\mathbf{f}(\mathbf{a}) = \mathbf{b} - \mathbf{a}$ が画素 P_a に与えられ、この2つのパッチ（あるいは、中心画素）のオフセットによるマッピングをニアレストネイバーフィールド（Nearest-Neighbor Field: NNF）という（図 7）。

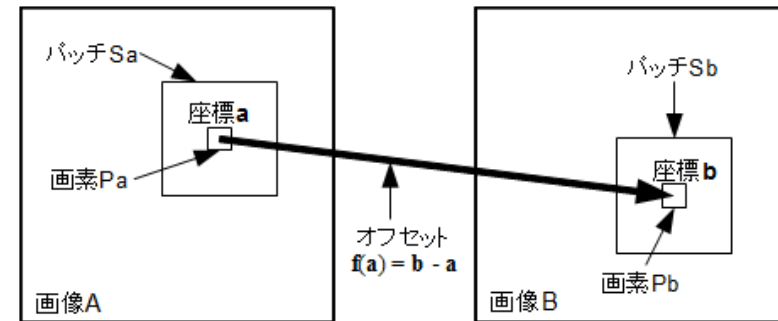


図 7 ニアレストネイバーフィールド

PatchMatch 法は、NNF のオフセットを初期化し、伝播とランダム探索の反復によって NNF を改良することで、近似的なニアレストネイバー（Approximate Nearest Neighbor: ANN）を高速に探索するアルゴリズムである（図 8）。まず、オフセットの初期化（図 8(1)）により、画像 A の画素ごとにランダムに画像 B の画素へのオフセットを与える。そして、以降、画像 A の画素ごとに走査線順で伝播とランダム探索を行う。伝播（図 8(2)）では、通常の走査線順（左から右、上から下）の場合、画像 A 上の現在の注目画素 P_a （座標 $\mathbf{a} = (x, y)$ ）に対して、すでに処理済みの上方画素（座標 $\mathbf{a}_y = (x, y-1)$ ）と左方画素（座標 $\mathbf{a}_x = (x-1, y)$ ）を考える。そして、画素 P_a のオフセットを $\mathbf{f}(\mathbf{a}) = (f_x(\mathbf{a}), f_y(\mathbf{a}))$ と表した場合、画像 B 上の3つの座標 $(x+f_x(\mathbf{a}), y+f_y(\mathbf{a}))$, $(x+f_x(\mathbf{a}_y), y+f_y(\mathbf{a}_y))$, $(x+f_x(\mathbf{a}_x), y+f_y(\mathbf{a}_x))$ の画素に対するパッチのうち、画像 A 上の画素 P_a に対するパッチ S_a に最も類似したものを求め、画素 P_a のオフセット $\mathbf{f}(\mathbf{a})$ を更新する。ランダム探索（図 8(3)）では、伝播によって更新した画素 P_a のオフセット $\mathbf{f}(\mathbf{a})$ に対する画像 B 上のパッチ S_b の周囲で、再度、類似パッチを探索する。まず、初期の探索範囲を、 S_b を中心とした最大探索半径 w ($1 < w \leq \max(\text{wide}, \text{height})$, wide と height は画像 B の横と縦の画素数) 以内の正方形領域とする。そして、この探索範囲内からパッチ S_b' をランダムに1つ取り出し、画像 A 上のパッチ S_a との類似度を求める。現在のオフセット $\mathbf{f}(\mathbf{a})$ に対するパッチ S_b よりも類似度が高ければオフセット $\mathbf{f}(\mathbf{a})$ を更新する。次に、比率 α ($0 < \alpha < 1$) で探索範囲を狭めて同様の処理を行ない、これを探索半径が1画素以下になるまで繰り返す。なお、2つのパッチ間の類似度は、パッチ内の画素ごとの RGB 値の差分の二乗の和で評価する。以上の走査線順の伝播とランダム探索は、オフセットの更新が無くなるまで反復する。その際、反復ごとに走査線順の方向を逆にする。すなわち、通常の走査線順（左から右、上から下）と逆方向の走査線順（右から左、下から上）を交互に繰り返す。

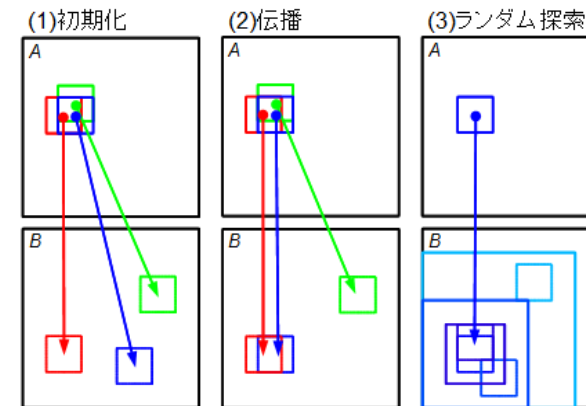


図 8 PatchMatch 法

4.2 ユーザ制御型テクスチャ合成法

ユーザ制御型テクスチャ合成法では、出力画像と同じ解像度の目的画像を用意することで、出力画像に目的画像の色分布を反映させることができる(図9)。この合成法では、出力画像の画素ごとの色を走査線順に決定していく。出力画像上の現在の注目画素 P_a の座標 a に対して、出力画像上の座標 a に対する左上 L 型パッチと目的画像上の座標 a に対する右下 L 型パッチを合わせた正方形パッチを考える。そして、この正方形パッチに類似したパッチを入力画像から探索し、その中心の画素の色を出力画像の注目画素 P_a に与える(図10)。

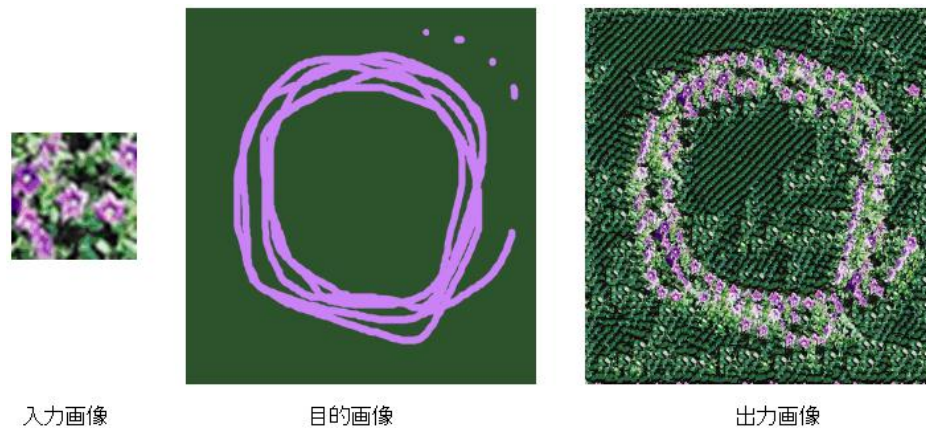


図9 ユーザ制御型テクスチャ合成法

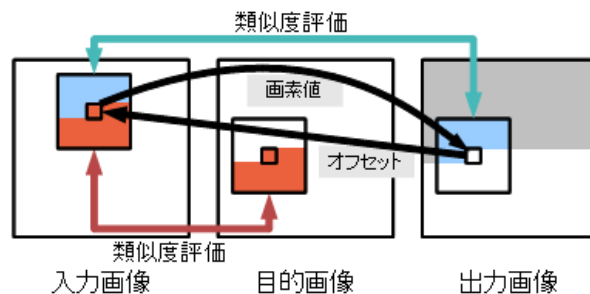


図10 ユーザ制御型テクスチャ合成法におけるパッチ間の類似度評価

5. テクスチャ合成による自由視点画像の品質改善

本研究のテクスチャ合成法は、PatchMatch法とユーザ制御型テクスチャ合成法を組み合わせるものになる。まず、3節の方法で生成したボクセルモデルを自由視点に対してレンダリングすることで“仮の”自由視点画像を生成する。先に述べたように、通常、ボクセルモデルの解像度や復元精度の問題などにより、この仮の自由視点画像は低品質となることが一般的である。そこで、本手法では、仮の自由視点画像を目的画像、カメラで撮影した元の参照画像を入力画像として、4.2節で述べたユーザ制御型テクスチャ合成を行うことで、その出力画像を高品質化した自由視点画像として得る。このとき、初期化修正反復と合成反復という2段階の反復処理により類似パッチの探索を行なうが、この探索方法には4.1節で述べたPatchMatch法を用いる。

まず、通常のPatchMatch法と同様に、出力画像の画素ごとにランダムに入力画像の画素へのオフセットを与えることでNNFの初期化(図11(1))を行う。次に、初期化修正反復(図11(2))では、出力画像の実際の合成に入る前に、初期化したNNFに対して目的画像を反映させるための修正を行う。この処理では、出力画像上で走査線順に座標 a の画素 P_a ごとに、PatchMatch法の伝播とランダム探索の反復により、目的画像上の座標 a の画素を中心としたパッチと類似するパッチを入力画像上で求め、得られたオフセットでNNFを更新する。このとき、同時に、求められた入力画像上の類似パッチの中心画素の色を出力画像の画素 P_a に与える。続いて、合成反復(図11(3))でも、初期化修正反復と同様のPatchMatch法による反復処理により、NNFと出力画像の画素の色を更新する。ただし、初期化修正反復とは類似度を評価するパッチが異なり、出力画像上の現在の注目画素 P_a の座標 a に対して、出力画像上の座標 a に対する左上 L 型パッチと目的画像上の座標 a に対する右下 L 型パッチを合わせた正方形パッチに対して入力画像上の類似パッチを求める。そして、最終的に得られる出力画像を高品質化された自由視点画像とする。

4.1節で述べたように、通常のPatchMatch法では、伝播において参照する画素は注目画素と上方画素、左方画素の3つであるが、本手法では、注目画素から参照する画素までの距離を可変にし、任意の距離 L 以内の画素を参照可能にした。これにより、入力画像のパターンを目的画像の色分布に合わせやすくなった。なお、上記の2段階の反復処理のいずれでも、終了条件は、各反復回におけるオフセットの更新率(= (その反復回にオフセットの更新がなされた画素数) / (初期化修正反復1回目でオフセットの更新がなされた画素数))が閾値 T 以下になる場合とした。これに加え、初期化修正反復では、反復回数が最大回数 M 回に達した場合にも反復を終了し、次の合成反復に移行することにした。また、合成反復では、反復ごとに走査線順の方向を逆にした場合、出力画像と目的画像のL型パッチの上下左右の向き

が逆になり、出力画像上の同じ座標に対する類似度の評価値が変化することから、オフセットの更新率が閾値 T 以下に収束しない現象が見られた。そこで、合成反復では、すべての反復で通常の走査線順（左から右、上から下）とした。

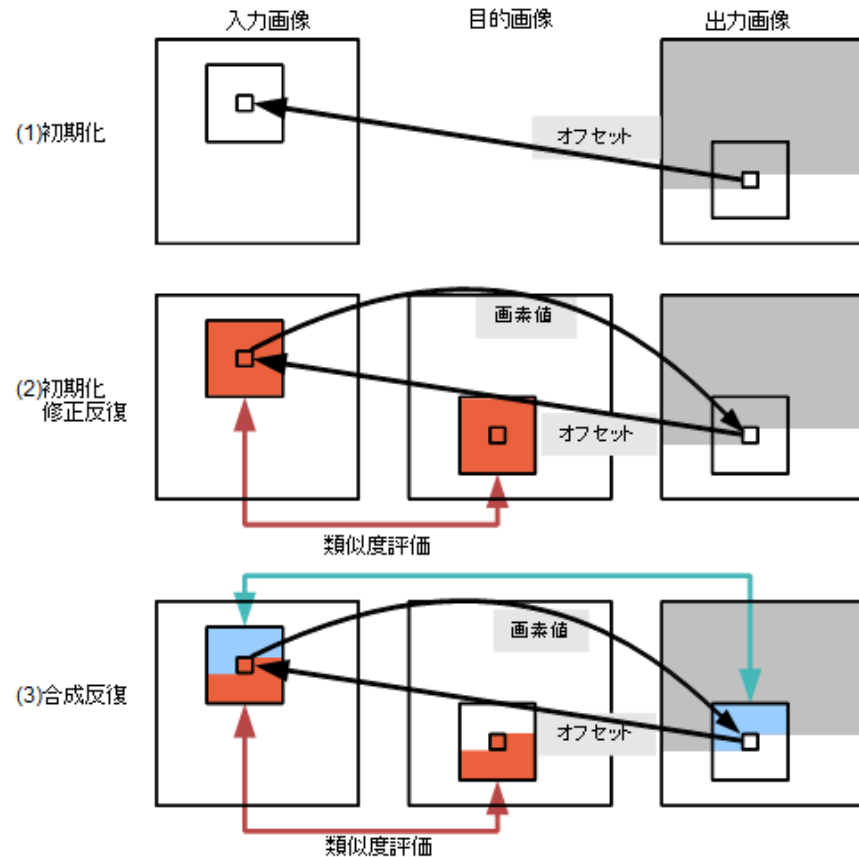


図 11 本手法のテクスチャ合成におけるパッチ間の類似度評価

6. 実験

本手法の有効性を検証するため、以下の実験を行った。青、黄、紫の花が生けてある鉢を対象物体とし、その周囲の等間隔の 8 方向からカメラで撮影して得られた参照画像を用いて、本手法による自由視点画像の生成を行なった。本実験では、ボクセル空間の解像度を 256^3 、参照画像と自由視点画像の画像サイズを 400×400 、テクスチャ合成のパッチサイズを 41×41 、伝播の際に参照する画素までの距離 L を 4、ランダム探索のときの最大探索半径 w を出力画像のサイズの $1/4$ (100×100)、比率 α を 0.8、反復の終了条件である閾値 T を 0.06、初期化修正反復の最大反復回数 M を 10 とした。これにより、探索時に参照するパッチの個数は、伝播の際は 17 個、ランダム探索の際は 21 個となる。図 12 に実験結果を示す。最上図の参照画像に対して、(a)から(e)は、参照画像を撮影したカメラ（参照カメラ）位置から隣のカメラ位置に向けて物体を中心に右回りに一定の角度 (5°) ずつ離れる位置に自由視点を設定して生成した自由視点画像を表す。それぞれ、左図はボクセルモデルをレンダリングした仮の自由視点画像、右図は本手法によるテクスチャ合成を適用した後の画像である。また、表 1 に計算時間と反復回数を示す。

図 12 から、低解像度のボクセルモデルをレンダリングしただけの低品質な画像に比べ、本手法によるテクスチャ合成を適用した画像では品質の改善がなされていることが分かる。特に、参照カメラとの角度差が小さいほど、参照画像の品質とほぼ変わらない結果が得られた。この場合、NNF の更新の反復回数は少なく、計算時間も短くなった。しかし、生成された画像の妥当性を考えた場合、特に、参照カメラとの角度差が大きいほど、品質は改善されてはいるものの、自由視点から見えるであろう実際の画像とは異なる、物体の見え方の妥当性の低い画像が得られることが多かった。この場合、NNF の更新の反復回数が多くなり、それに比例して計算時間も長くなる傾向にあり、オフセットの更新率が収束しないケースもあった。この問題の大きな原因の一つは、ボクセル空間の解像度の低さであると考えられる。また、それに加え、カメラ校正の誤差等によるボクセルモデルの復元精度の低さも挙げられる。本研究は、参照画像の枚数が少なく、かつ、ボクセルモデルが低解像度である場合でも高品質な自由視点画像を生成することが目的であるが、本実験により、3 節の方法で復元したボクセルモデルをそのままレンダリングしただけの画像では、自由視点がカメラ位置から離れるにつれて、自由視点に対する物体の見え方の妥当性が失われ、テクスチャ合成の目的画像として十分な役割を果たせなくなる傾向にあることが分かった。この改善が今後の課題である。



参照画像



(b) 参照カメラとの角度差 : 5°



(a) 参照カメラとの角度差 : 0°



(c) 参照カメラとの角度差 : 10°



表 1 計算時間と反復回数. 単位は分. カッコ内は反復回数.

角度差	初期化 修正反復	合成反復	合計
0°	4.44(3)	3.06(2)	7.50 (5)
5°	10.42(7)	7.65(5)	18.08(12)
10°	10.36(7)	21.57(14)	31.93(21)
15°	14.57(10)	25.65(17)	40.23(27)
20°	14.52(10)	31.59(21)	46.15(31)



(d) 参照カメラとの角度差 : 15°



(e) 参照カメラとの角度差 : 20°

図 12 参照画像と異なる自由視点からの自由視点画像.

7. おわりに

本研究では, 多視点画像を用いて復元したボクセルモデルを描画した自由視点画像に対し, PatchMatch 法を利用したユーザ制御型テクスチャ合成法により品質の改善を行う手法を提案した. 実験により, 本手法による自由視点画像の品質の改善が確認でき, 特に, 参照カメラに近い自由視点に対して, 高品質な自由視点画像を得ることができた. しかし, 参照カメラから遠い自由視点に対しては, 物体の見え方の妥当性の低い自由視点画像が生成される傾向があり, この改善が今後の課題である. また, 近隣の自由視点画像間での連続性を考慮することも課題であり, NNF の初期化に近隣の自由視点画像の合成後の NNF を用いる手法の検討などを予定している.

謝辞 本研究を実施するにあたり, 様々な貴重なご意見を頂いた岩手大学工学部千葉則茂教授に感謝致します. 本研究の一部は科学研究費補助金 (基盤研究 (C) 21500090) の援助を受けている.

参考文献

- 1) G. Slabaugh, B. Culbertson, T. Malzbender, R. Shafer, A Survey of Methods for Volumetric Scene Reconstruction from Photographs, Proceedings of the International Workshop on Volume Graphics 2001, pp.81-100, 2001.
- 2) M. Ashikhmin, Synthesizing Natural Textures, ACM Symposium on Interactive 3D Graphics 2001, pp.217-226, 2001.
- 3) 高橋実, 藤本忠博, 原美オサマ, 千葉則茂, テクスチャ合成を用いた複数カメラ画像からの高品質ボクセルモデルの構築法, 第 25 回 NICOGRAPH 論文コンテスト, pp.S6P1-1-8, 2009 年.
- 4) S. M. Seitz, C. R. Dyer, Photorealistic Scene Reconstruction by Voxel Coloring, Proceedings of Computer Vision and Pattern Recognition Conference, pp.1067-1073, 1997.
- 5) C. Barnes, E. Shechtman, A. Finkelstein, D. B. Goldman, PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing, Proceedings of SIGGRAPH 2009, pp.24(1)-(12), 2009.