

LJ-001

イメージフュージョンによる任意視点画像の鮮鋭化手法 Post-Rendering View Fusion for Non-Aliased Light Field Rendering

久保田 彰[†]
Akira Kubota

相澤 清晴[‡]
Kiyoharu Aizawa

1. まえがき

多視点から撮像した複数枚の画像を用いて、任意視点からの画像を生成する Image-Based Rendering (IBR) 手法 [1, 2] が数多く提案されている。IBR 手法は次の二つの手法に大きく分けられる。一つは、多視点画像から対象シーンの 3 次元構造を推定し、得られた 3 次元モデルに基づいてレンダリングを行う手法である。この手法は Image-Based Modeling and Redering (IBMR) 手法とも呼ばれている。IBMR 手法は、その構成上、シーンの 3 次元構造の推定精度が生成画像の品質に大きく影響する。対象シーンが複雑な実シーンの場合、一般に、その 3 次元構造を高精度に推定することは困難であるため、視点移動に伴って、生成画像に視覚的な歪みが生じる問題がある。

もう一つは、密な間隔で取得された多視点画像の補間に基づいた手法である。この手法はシーンの 3 次元構造の推定が不要であるという特長がある。たとえば、Light Field Rendering (LFR) [3] は、対象シーンを 1 枚の平面 (focal plane) として近似し、任意視点画像をレンダリングする。しかしながら、十分な精度でレンダリングできるシーンの奥行きが非常に狭いという問題がある [4]。逆に言えば、奥行きのあるシーンをレンダリングするためには、高密度に取得された多視点画像が必要となる。取得密度が十分でない場合、補間に用いられる画素値 (あるいは光線) の対応がずれるため、focal plane から離れたシーンにボケやゴーストといった劣化が生じる。この現象は、光線のサンプリング理論の観点から、エイリアジングとしても解析されている [5]。

近年、図 1 に示すように、次の 2 段階からなる、IBR の新しい枠組みが検討されている。まず、複数枚の focal plane を設定し、それらに基づいて LFR により複数枚の任意視点画像を生成する。この段階で得られた各画像では、focal plane 近傍の領域は鮮鋭に生成されるが、focal plane から離れた領域には劣化が生じる。次に、これらの画像を統合することによって、劣化を大きく低減した高品質な画像を生成する。IBMR とは異なり、レンダリング後の複数の画像を統合する構成から、本手法を Image-Based Rendering and Fusing (IBRF) と呼ぶことにする。画像の統合手法として、画像中の劣化のない鮮鋭な領域を検出し、それらを統合する方法が提案されている。鮮鋭領域の検出には、光線の色的一致を評価する方法 [6] や LFR に特化した合焦判定法 [7] が提案されている。これらの手法は、結果的に所望の視点位置に依存した奥行きマップを推定している。

本稿では、IBRF の枠組みにおける画像統合に対して

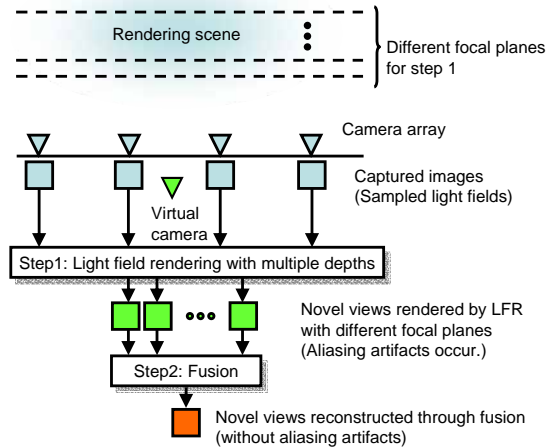


図 1: Image-Based Redering and Fusing (IBRF) の概観。本稿では、Step2 において新しい画像統合手法を提案する。

新しい手法を提案する。提案手法では、奥行きごとに定義されたテクスチャの線形和によって画像をモデル化し、画像統合を連立作用素方程式を解く問題として定式化する。この方程式を反復的に解くことによって、局所的な鮮鋭領域の検出や領域分割を行うことなく、効果的に鮮鋭領域を統合を行うことができる。

2. 提案手法

2.1 画像のモデル

(a) 奥行きレイヤによるシーン構造の表現

仮定する focal plane の枚数を N 、その奥行きを z_n ($n = 1, 2, \dots, N$) とする。図 2 に示すように、与えられた視点位置から見た対象シーンを奥行き z_n にある複数の奥行きレイヤ (平面) で近似する。仮想カメラで観測される奥行きレイヤの画像を depth texture と呼び、奥行き z_n に対する depth texture を次のように定義する。

$$f_n(x, y) \stackrel{\text{def.}}{=} \begin{cases} f(x, y), & \text{if } d(x, y) = z_n \\ 0, & \text{otherwise} \end{cases}, \quad (1)$$

$(n = 1, 2, \dots, N).$

ここに、 (x, y) は画素の座標、 $f(x, y)$ は所望の鮮鋭画像を表し、 $d(x, y)$ は解像度が N である奥行きマップに相当する。すなわち、 f_n は与えられた視点位置における奥行きマップに基づいて、所望の鮮鋭画像を各奥行きごとに領域分割したものとなる。ただし、奥行きマップは未知であるため、すべての depth texture は未知である。

(b) 線形和による画像のモデル化

所望の鮮鋭画像 f は、depth texture の和として次の

[†]神奈川大学工学研究科 ハイテク・リサーチ・センター, High Tech. Research Center, School of Eng., Kanagawa Univ.

[‡]東京大学大学院新領域創成科学研究科, School of Frontier Science, the Univ. of Tokyo

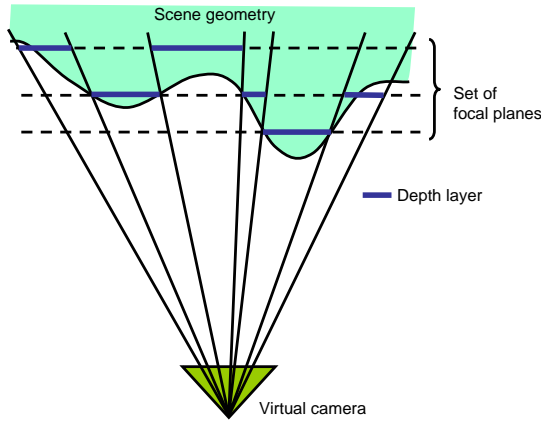


図 2: 奥行きレイヤによるシーン構造の近似 .

ようにモデル化できる .

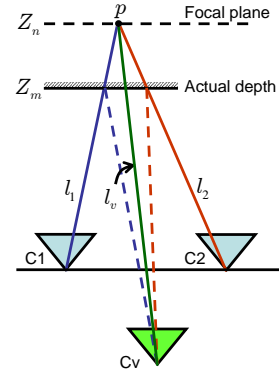
$$f = f_1 + f_2 + \dots + f_N. \quad (2)$$

奥行き z_n の focal plane に基づいて LFR によりレンダリングされた任意視点画像を g_n とする . focal plane と同じ奥行きにある f_n 以外の depth texture に劣化が生じるため , これらの画像を次のようにモデル化する .

$$\begin{cases} g_1 = f_1 + h_{12}f_2 + h_{13}f_3 + \dots + h_{1N}f_N \\ g_2 = h_{21}f_1 + f_2 + h_{23}f_3 + \dots + h_{2N}f_N \\ \vdots \\ g_N = h_{N1}f_1 + \dots + h_{NN-1}f_{N-1} + f_N. \end{cases} \quad (3)$$

h_{nm} は , focal plane の奥行きが z_n のときに奥行き z_m の depth texture f_m に生じる劣化を表す作用素である .
(c) 劣化過程のフィルタモデル

劣化作用素 h_{nm} は , focal plane と depth texture の奥行きの違いによって生じる「対応点のずれ」として考えることができる . 図 3 を用いて , これを簡単に説明する . 簡単のため , 直線に配置された 2 台のカメラ C1 と C2 で取得された画像 (光線 l_1 と l_2) から , 仮想カメラ Cv の画像 (光線 l_v) を LFR によって生成する場合を考える . シーン中の奥行き z_m の奥行きレイヤに対して , 奥行き z_n の focal plane を設定した場合 , 光線 l_v は focal plane 上の点 p に対応した 2 本の光線 l_1 と l_2 の重み和として生成される . しかしながら , 実際の奥行きレイヤは focal plane とは異なる奥行きにあるため , 異なる点からの光線 (図中の 2 本の破線) が l_v に合成されることになる . この「対応点のずれ」は , 2 次元配置のカメラを用いた場合 , 4 つのインパルスからなるフィルタとしてモデル化できることを示している . ただし , ここではシーンがランバート反射特性であり , オクルージョンの影響が無視できるものとする . このフィルタは , 視点位置や focal plane の奥行きだけでなく , 任意視点画像の画素の座標にも依存するため , 空間可変 (space variant) となる . これは , 特定の奥行き領域で h_{nm} がモデル化されているという意味ではない . h_{nm} は画像全体への作用素である .


 図 3: 劣化過程 h_{nm} の説明 .

なお , 一般のカメラで取得された多焦点画像の統合手法 [8] において , 式 (3) と同様な画像のモデル化が利用されているが , 焦点ボケの場合 , h_{nm} は空間不変な低域通過フィルタとなる . この劣化特性の違いが , 従来の Depth From Focus 手法 [9] とは大きく異なる点である .

2.2 反復再構成

本稿で扱う画像統合は , 与えられた g_n および h_{nm} から , f を求める問題として定式化される . 式 (3) は , N 元の連立作用素方程式となっている . そこで , 式 (3) から個々の f_n を解き , その和を求める .

本稿では , 一般の連立方程式の解法である Gauss-Seidel 法 [10] に基づいて , 以下のように f_n を求める . まず , 式 (3) において f_n と g_n の項をそれぞれ左辺と右辺に移項し , 次の式を得る .

$$\begin{cases} f_1 = g_1 - h_{12}f_2 - h_{13}f_3 - \dots - h_{1N}f_N \\ f_2 = g_2 - h_{21}f_1 - h_{23}f_3 - \dots - h_{2N}f_N \\ \vdots \\ f_N = g_N - h_{N1}f_1 - h_{N2}f_2 - \dots - h_{NN-1}f_{N-1}. \end{cases} \quad (4)$$

つぎに , 解の初期値として , $\{f_1^{(0)}, f_2^{(0)}, \dots, f_N^{(0)}\}$ を設定し , これらを式 (4) の第 1 番目の式の右辺へ代入する . これにより , f_1 の 1 回目の反復解 $f_1^{(1)}$ を次のように得る .

$$f_1^{(1)} = g_1 - h_{12}f_2^{(0)} - h_{13}f_3^{(0)} - \dots - h_{1N}f_N^{(0)}. \quad (5)$$

このとき , f_1 の初期解 $f_1^{(0)}$ は $f_1^{(1)}$ の結果に寄与しない . 続いて , ここで得られた $f_1^{(1)}$ と $f_3^{(0)}, \dots, f_N^{(0)}$ を式 (4) の第 2 番目の式の右辺へ代入することにより , f_2 の 1 回目の反復解 $f_2^{(1)}$ を次のように得る .

$$f_2^{(1)} = g_2 - h_{21}f_1^{(1)} - h_{23}f_3^{(0)} - \dots - h_{2N}f_N^{(0)}. \quad (6)$$

以下 , 同様にして得られた解をすぐに次の式へ順次代入することによって , 反復解を得る . k 回目の反復解 $f_n^{(k)}$ は次の式のように表される .

$$\begin{cases} f_1^{(k)} = g_1 - h_{12}f_2^{(k-1)} - h_{13}f_3^{(k-1)} - \dots - h_{1N}f_N^{(k-1)} \\ f_2^{(k)} = g_2 - h_{21}f_1^{(k)} - h_{23}f_3^{(k-1)} - \dots - h_{2N}f_N^{(k-1)} \\ \vdots \\ f_N^{(k)} = g_N - h_{N1}f_1^{(k)} - h_{N2}f_2^{(k)} - \dots - h_{NN-1}f_{N-1}^{(k)}. \end{cases} \quad (7)$$

k 回目の f の反復解 $f^{(k)}$ はこれらの和, $f_1^{(k)} + f_2^{(k)} + \dots + f_N^{(k)}$ として求める. 次章の実験結果で示されるように, 本手法により効果的に各任意視点画像 g_n の鮮鋭領域が f に統合される.

この反復法では, 局所的な領域分割や奥行き推定などの処理が行われていない. 行われている処理は, 劣化作用素 h_{nm} のフィルタリングと加減算だけである. 前章(c)において述べたように, h_{nm} のフィルタリングは特定の領域で行われるのではなく, 画像全体で行われる. 領域を検出する機能はない.

3. 実験

実写の多視点画像を用いて, 提案手法による任意視点画像の生成を行った. 多視点画像には, 「筑波大学多視点画像データベース」を使用した. それらの画像は, カメラを水平および垂直方向に 20 [mm] 間隔で移動させて, 合計 81 視点から取得したものである. 画像の解像度は 480×360 [pixels] とした. 対象シーンは人形であり, その奥行き範囲は 590–800 [mm] である. 本実験では, この範囲をレンダリングの対象とした. この撮像条件では, 隣接した多視点画像間での最大および最小の視差は, 約 36 および 26 [pixels] となる. その差は 10 [pixels] である. これは, Plenoptic sampling [5] の理論から, LFR により劣化の生じない任意視点画像を得るためには双方向に 5 倍ずつの台数のカメラ (すなわち, カメラ間隔 4[mm] で合計約 2,000 台) が必要であることを意味している. この条件で画像を取得するのは現実的ではないことがわかる.

図 4 に 5 枚の focal plane に基づいて LFR により生成した 5 枚の任意視点画像 g_n ($n = 1, \dots, 5$), それらから提案手法によって求めた 5 回目の反復解 $f_n^{(5)}$ および最終的な鮮鋭画像 $f^{(5)}$ を示す. $f_n^{(5)}$ に関しては, 見やすくするため明度およびコントラストを調節している. 各 focal plane の距離は, 対象の奥行き範囲を視差空間が等分されるように設定した. 解の初期値は $f_n^{(0)} = g_n/5$ ($n = 1, \dots, 5$) とした. 図 4 において, g_n にはボケやゴーストの劣化が見られるが, $f^{(5)}$ には g_n における鮮鋭領域が良好に統合されて, 全体で鮮鋭な画像となっている. ステップエッジ付近では, 若干の劣化が見られるが, オクルージョンの影響は視覚的に大きく生じていない. 得られた反復解 $f_n^{(5)}$ は, 式 (2) で定義した depth texture, すなわち領域分割された奥行きごとのテクスチャとなっていない. しかしながら, それらの和である $f^{(5)}$ は良好な鮮鋭画像となっている. これについては, 本手法の誤差の収束性として次章で考察する.

4. 収束性の考察

任意の視点位置, focal plane の枚数および奥行き条件下で, 本手法の収束を数学的に証明することは一般的に難しい. そこで, $N = 3$ において誤差の収束をシミュレーションすることによって, 本手法の収束性を示す.

$f_n^{(k)}$ と $f^{(k)}$ の誤差としてそれぞれ $e_n^{(k)} = f_n^{(k)} - f_n$ と $e^{(k)} = f^{(k)} - f$ を定義する ($n = 1, 2, 3$). このとき, $e^{(k)} = e_1^{(k)} + e_2^{(k)} + e_3^{(k)}$ が成り立っている. 初期誤差として $e_n^{(0)}$ に -128 から 128 の範囲のランダム信号を与え,

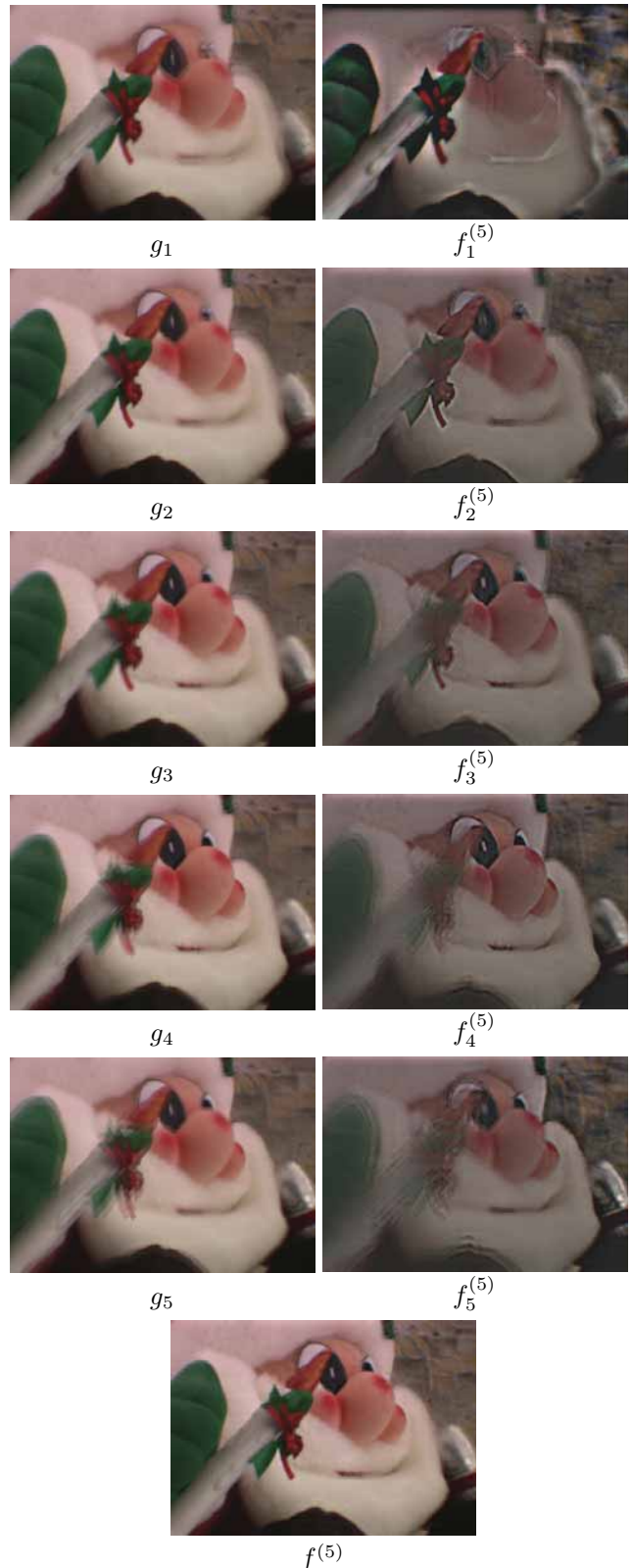


図 4: 実験結果. g_n は異なる奥行きに設定した focal plane に基づいて LFR によって生成された任意視点画像, $f_n^{(5)}$ および $f^{(5)}$ は提案手法によって求められた 5 回目の反復解である ($n = 1, \dots, 5$).

反復 10 回目の誤差 $e_n^{(10)}$ および $e^{(10)}$ を計算した。水平 1 ラインの結果を図 5 に示す。細い線は初期誤差，太い線は反復 10 回目の誤差を表す。各 $e_n^{(10)}$ は，初期誤差に比べて減衰しているが，十分に小さくなっていない。しかし，その和である $e^{(10)}$ は，それらの誤差が打ち消しあって，十分に小さく減衰していることがわかる。また，この結果は，任意の初期値に対して本手法が収束することを示唆している。

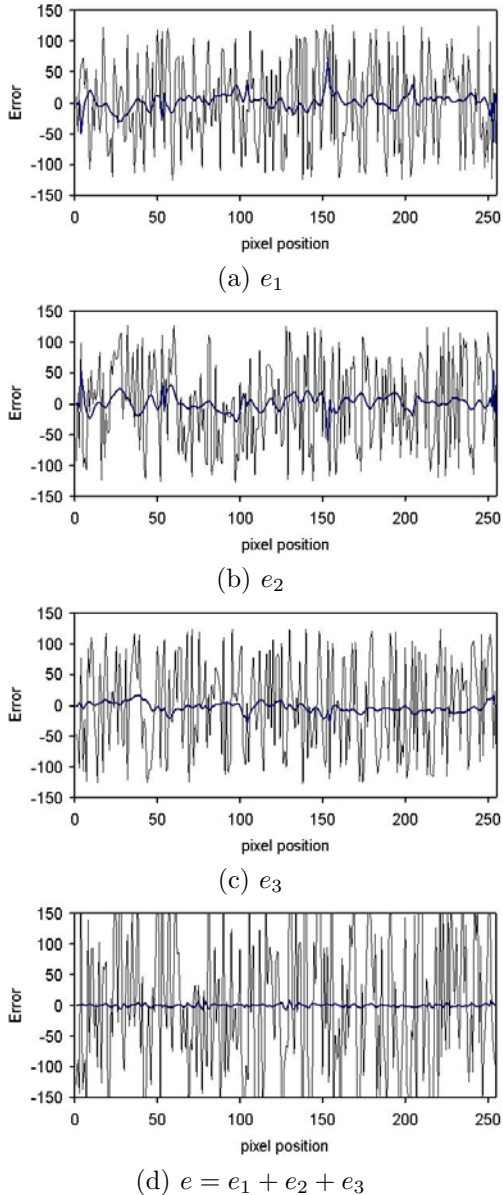


図 5: 誤差の収束のシミュレーション結果。細い線は初期誤差，太い線は反復 10 回目の誤差を表す。

5. むすび

本稿では，複数枚の focal plane に基づいて Light Field Rendering によりレンダリングした複数枚の任意視点画像を統合することによって，対応点ずれによる劣化が低減された鮮鋭画像を生成する手法を提案した。提案手法は，複数枚の任意視点画像を奥行きごとの画像の線形和としてモデル化し，連立作用素方程式を反復的に解くことによって，効果的に所望の画像の生成を行うことができる。その際，局所的な領域分割や奥行き推定をする必要がないという特長がある。実写画像を用いた実験により，本手法の有効性を示し，シミュレーションにより誤差の収束性を示した。

今後の課題として，本手法の収束性に関する数学的な考察，オクルージョン領域の誤差解析やテクスチャマッピングによる手法の高速化などに取り組む予定である。

謝辞：本手法に関して，有益なご意見を頂いた Carnegie Mellon University 教授 Tsuhan Chen 氏に感謝致します。

参考文献

- [1] H-Y. Shum, S. B. He, and S-C. Chan, "Survey of Image-Based Representations and Compression Techniques," *IEEE Trans. on CSVT*, 13, 11, pp. 1020–1037, 2003
- [2] C. Zhang and T. Chen "A Survey on Image-Based Rendering – Representation, Sampling and Compression," *EURASIP Signal Processing: Image Communication*, 19, 1, pp. 1–28, 2004
- [3] M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH96*, pp.31–42, 1996
- [4] 高橋，苗村，原島，"自由視点画像合成における光線情報のサンプリングと被写界深度に関する理論"，*映情学誌*，57, 10, pp. 1292–1299, 2003
- [5] Jin-Xiang Chai, X. Tong, S.-C. Chany and H.-Y. Shum, "Plenoptic Sampling," *SIGGRAPH2000*, pp.307–318, 2000
- [6] A. Isaksen, L. McMillan, S. J. Gortler "Dynamically Reparameterized Light Fields," MIT-LCS-TR-778, 1999
- [7] K. Takahashi, A. Kubota, T. Naemura "All in-Focus View Synthesis from Under-Sampled Light Fields," *Proc. VRSJ ICAT*, pp. 249–256, 2003
- [8] 久保田，相澤，"線形フィルタによる 2 枚の合焦画像からの任意ぼけ画像の再構成"，*信学誌*，J83-D-II，12，pp. 2663–2674，2000
- [9] S. K. Nayar, "Shape from Focus System," *CVPR92*, pp. 302–308, 1992
- [10] David M. Young, *Iterative Solution of Large Linear System*, Academic Press 1971