

博士学位論文

多眼画像を用いた全焦点自由視点画像の
実時間合成に関する研究

2005 年 12 月 16 日

指導教員 原島 博 教授

東京大学大学院情報理工学系研究科
電子情報学専攻 37413

高橋 桂太

Abstract

In recent years, visual media technologies have made a great progress with the progress in computer power. It is not a dream that we can deal with the complete visual information of the real 3-D world in which we live, as well as 2-D images of that. Related technical fields range in 3-D digital imaging, computer vision, computer graphics, and multi-dimensional signal processing. The technologies to process 3-D visual information are expected in many application fields such as telecommunication, virtual reality, 3-D contents production, and so on, and will increase their importance in the future.

As one of the methods to represent 3-D visual information of the real world, this paper focuses on free-viewpoint image synthesis using multi-viewpoint images. Generally speaking, this technology consists of two steps, analysis and synthesis. The analysis step is a problem of "image-based modeling" in the computer vision field. In this step, multi-viewpoint images are analyzed to estimate the structure of the scene, such as geometrical shapes. Typically, this step is computationally expensive, and requires significant off-line processing. The synthesis process is a problem of "image-based rendering" in the computer graphics field. This step calculates how the scene looks from the given viewpoints to synthesize images, using structure data estimated through the precedent analysis step as priori knowledge. The eventual goal is to achieve real-time processing of 3-D dynamic scenes, in which multi-viewpoint videos are captured, processed, and converted into free-viewpoint videos in real-time. For this purpose, an integrated discussion beyond the boundary between the above two steps is required to optimize the entire algorithm. The goal of this paper is to synthesize all in-focus free-viewpoint images in real-time from multi-view images captured with an array of cameras. Note that this paper assumes there is no defocus blurring on the input images themselves, and uses the word "focus" for the phenomena on the synthesized images.

One of the main contributions of this paper is a study of real-time image-synthesis methods. I propose two methods: one is based on the optimized multi-view stereo matching, the other one is based on my original scheme for depth estimation, which is named the focus measurement. The method based on the focus measurement arises from the extension of light field rendering, which assumes the entire scene could be approximated with a single plane. This method consists of two steps for a certain desired viewpoint. First, several images are synthesized using light field rendering with different assumed depths. In these images, focus-like effects are caused according to the assumed depth, and the out-of-focus parts become unclear and ghosted. Then,

the in-focus parts of those images are detected by the focus measurement and integrated into one final image, resulting an all in-focus image which is clear throughout the scene. Both of the proposed methods can be implemented efficiently using graphics hardware accelerations, and can achieve real-time rendering with stored multi-viewpoint image datasets. The latter method is also applied to a physical system with a lens-array to confirm that dynamic 3D scenes can be rendered in real-time with this method.

Another important contribution of this paper is a theoretical systemization. First, this paper introduces a theory of the depth-of-field, which represents the quality limitation of the images synthesized with a single plane model to approximate the scene. Then, this paper studies how to optimize multi-view stereo matching for real-time free-viewpoint image synthesis. This optimization reveals that the matching problem should be directly solved in the coordinate system that is centered at the desired viewpoint, and is applied to the first proposed method for image synthesis. In the second proposed method, a new scheme named the focus measurement is introduced. The spatial domain analysis of this scheme provide a practical algorithm which can be applied to inputs from arbitrary camera arrays. The frequency domain analysis gives a unique theoretical grounding based on a theory of multi-dimensional signal processing. The significance of this scheme is that it introduces a novel approach to the depth estimation problem for free-viewpoint image synthesis, in which the synthesized images, instead of the input images, are directly analyzed. In addition, a generalized framework provided by this scheme includes multi-view stereo matching method.

The methods proposed in this paper realize free-viewpoint image synthesis in real-time with high quality, and will be more refined and applied for real time applications such as telecommunication. The theoretical studies in this paper try to approach to the problem of free-viewpoint image synthesis from the three technical fields, computer graphics, computer vision, and signal processing, beyond their boundaries. I hope this approach indicates the way to a new research field of 3-D visual information processing, in which researches in related areas would be compared, related and integrated with each other.

目次

第 1 章	序論	1
1.1	本研究の目的	2
1.2	本論文の構成	3
第 2 章	研究の背景	6
2.1	はじめに	7
2.2	三次元統合画像通信と光線空間法	7
2.3	イメージベースレンダリングについての概論	8
2.3.1	幾何情報を用いない手法	9
2.3.2	暗示的に幾何情報を用いる手法	12
2.3.3	明示的な幾何情報を用いる手法	13
2.3.4	画像情報と幾何情報のトレードオフ関係	15
2.4	動空間を対象とした自由視点画像合成システムの例	16
2.4.1	オブジェクトを対象とするシステム	16
2.4.2	シーンを対象とするシステム	19
2.5	多次元信号処理から見た自由視点画像合成	21
2.5.1	Plenoptic Sampling [53]	22
2.5.2	Generalized Plenoptic Sampling [98]	24
2.5.3	再構成フィルタと合成画像	26
2.6	本研究の位置づけ	27
	参考文献	28
第 3 章	自由視点画像合成におけるサンプリングと被写界深度の理論	36
3.1	まえがき	37
3.2	関連研究	37
3.2.1	Light field rendering のアルゴリズム	38
3.2.2	Plenoptic Sampling	39
3.2.3	奥行き範囲に対する定式化	40
3.3	Light Field Rendering における被写界深度の提案	41
3.3.1	導出	41
3.3.2	3 通りの表現	41

3.3.3 レンダリングカメラの解像度についての考察	43
3.4 他の理論との関係	43
3.4.1 関連理論の導出	43
3.4.2 各理論の比較	45
3.5 実験	46
3.6 むすび	47
参考文献	50
3.7 付録	51
3.7.1 light field rendering における被写界深度	51
3.7.2 等価被写界深度	51
3.7.3 定数倍の項についての説明	52
3.7.4 平行移動の項についての説明	53
第 4 章 全焦点自由視点画像合成のための多眼ステレオ法の最適化	55
4.1 はじめに	56
4.2 関連研究	57
4.2.1 自由視点画像合成と形状モデル推定	57
4.2.2 視点依存形状モデル	58
4.3 提案手法のアルゴリズム	60
4.3.1 設定	60
4.3.2 レイヤモデル	61
4.3.3 局所的な color-consistency の評価	61
4.3.4 評価値画像空間での平滑化	63
4.3.5 最小値探索による奥行き決定	63
4.3.6 各画素の色づけ	64
4.3.7 擬似コード	64
4.4 実験	65
4.4.1 実装	65
4.4.2 処理時間	65
4.4.3 奥行き推定および画像合成の結果	68
4.5 むすび	70
参考文献	70
第 5 章 合焦判定法に基づく全焦点自由視点画像合成 – 空間領域における検討	73
5.1 はじめに	74
5.2 背景	75
5.3 提案アルゴリズム	76
5.3.1 各レイヤに対するレンダリング	78

5.3.2	“合焦尺度”の計算	78
5.3.3	“全焦点”画像の合成	80
5.3.4	奥行き情報の平滑化	80
5.4	実装と実験	81
5.4.1	実装	81
5.4.2	実験	83
5.5	むすび	87
	参考文献	88
第 6 章	合焦判定に基づく全焦点自由視点画像合成 – 周波数領域における検討	92
6.1	まえがき	93
6.2	背景	93
6.2.1	Light Field Rendering における “合焦判定” 法	93
6.2.2	他の奥行き推定法との比較	94
6.2.3	Light field の周波数解析と “焦点”	95
6.3	提案手法	97
6.3.1	Light Field Rendering における合焦判定の原理	97
6.3.2	“合焦” 判定による奥行き推定と “全焦点” 画像合成	99
6.4	実験	100
6.4.1	実写画像を用いた実験	100
6.4.2	定量的な評価	102
6.5	むすび	106
	参考文献	107
6.6	付録	108
6.6.1	再構成フィルタの空間波形の導出	108
6.6.2	近似的な実装	109
第 7 章	合焦判定に基づく全焦点自由視点画像合成法の実システムへの適用	111
7.1	まえがき	112
7.2	従来研究	112
7.2.1	LIFLET の撮像系	112
7.2.2	合焦判定に基づく全焦点画像合成	113
7.2.3	レンズアレイ画像を用いた自由視点画像合成	113
7.3	レンズアレイ方式自由視点画像合成における “合焦判定” 法の適用	115
7.3.1	奥行き制御レンズを考慮した light field rendering	116
7.3.2	focal plane の設定の最適化	118
7.4	実時間合成のための実装および実験	120
7.4.1	実時間合成	120

7.4.2 従来手法との画質の比較	121
7.5 むすび	124
参考文献	124
第 8 章 多次元信号処理から見た合焦判定法と多眼ステレオ法	127
8.1 まえがき	128
8.2 背景	128
8.3 light field rendering の周波数解析	129
8.3.1 問題設定	129
8.3.2 周波数解析	129
8.3.3 再構成フィルタと焦点との関係	130
8.4 周波数領域における合焦判定法	131
8.4.1 前提	131
8.4.2 light field におけるサブサンプリング	132
8.4.3 合焦判定の基本原則	133
8.5 合焦判定の空間領域への展開とステレオマッチング法の導出	134
8.5.1 空間領域への展開	134
8.5.2 再構成フィルタの空間波形と変数分離	135
8.5.3 画像合成の実装	136
8.5.4 ステレオマッチング法の導出	137
8.6 むすび	138
参考文献	138
8.7 付録	140
8.7.1 サブサンプリングされた信号の周波数特性	140
8.7.2 再構成フィルタの空間波形の導出	141
8.7.3 画像合成の実装についての数学的詳細	142
第 9 章 結論	144
9.1 本論文の成果	145
9.1.1 多眼画像を用いた全焦点自由視点画像の実時間合成の実践	145
9.1.2 自由視点画像合成に関する理論的体系化	146
9.2 今後の課題と展望	147
9.2.1 実システムへの応用	147
9.2.2 理論面の発展	148
参考文献	149
発表文献	150
研究論文 (査読付き論文)	150

国際会議論文	150
シンポジウム・研究会・大会等発表	151
謝辞	154

目次

1.1	本論文の構成	4
2.1	三次元統合画像通信の概念図 (文献 [1] より引用)	8
2.2	Light field のデータ形式とレンダリング (文献 [18] より引用)	11
2.3	Concentric mosaics (文献 [18] より引用)	12
2.4	View morphing (文献 [32] より引用)	13
2.5	MCOP の撮影 (左上), 構築 (右上) および再投影 (下) (文献 [41] より引用)	14
2.6	Surface light field (文献 [46] より引用)	14
2.7	Minimum sampling curve (文献 [53] より引用)	15
2.8	Virtualized Reality システム (文献 [61] より引用)	17
2.9	Image-based visual hulls (文献 [68] より引用)	17
2.10	大規模空間を対象とした三次元ビデオシステムの概念図 (左) と多層平面に基づく スケーラブルな三次元モデル (右) (文献 [76] より引用)	18
2.11	Video based rendering システム (文献 [79] より引用)	19
2.12	Yang らのカメラアレイシステム (左) と合成画像の例 (右) (文献 [22] より引用)	20
2.13	Self-reconfigurable camera array (文献 [89] より引用)	21
2.14	LIFLET システム (文献 [91] より引用)	21
2.15	Wilburn らのカメラアレイ: 用途に応じて様々な構成を取る (文献 [95] より引用)	22
2.16	パラメータ設定	23
2.17	光線信号の解析	23
2.18	Plenoptic sampling 理論	23
2.19	非ランバート成分を含む光線情報 (文献 [98] より引用)	25
2.20	オクルージョンを含む光線情報 (文献 [98] より引用)	25
2.21	左: 空間領域におけるサンプリング格子の構造, 右: 周波数領域における繰り返 しの格子の構造 (文献 [98] より引用)	26
2.22	スペクトラムサポートの最密充填 (文献 [98] より引用)	26
2.23	再構成フィルタの形状	27
2.24	再構成フィルタと合成画像 (文献 [100] より引用)	27
3.1	パラメータの取り方	38
3.2	等価被写界深度 (文献 [6] より引用)	40

3.3	多視点カメラシステムの解像度（文献 [7] より引用）	40
3.4	Light field rendering における被写界深度：縦軸は camera plane からの距離，横軸はそれぞれ (a) カメラ間隔，(b) 空間周波数の最大値，(c) focal plane の奥行き	42
3.5	提案理論からの等価被写界深度および多視点カメラシステムの解像度の導出	44
3.6	実験条件	46
3.7	画像合成：(a) カメラ間隔 $\Delta s = \Delta s_{\max}$ の場合，(b) $\Delta s > \Delta s_{\max}$ の場合 (c) 画像 (a) と画像 (b) の差分	48
3.8	Light field rendering における被写界深度とカメラ間隔についての実験：上の図は合成画像の PSNR，下の図は提案理論に基づいて計算された被写界深度を示している	49
3.9	等価被写界深度の定式化	51
3.10	平行移動の項についての説明	53
4.1	基本的な配置	60
4.2	color-consistency の評価	61
4.3	擬似コード	66
4.4	入力画像データ（「筑波大学多視点画像データベース」より）	67
4.5	City 画像群を用いた合成画像（上段）と対応する奥行きマップ（下段）	69
4.6	Plants 画像群を用いた合成画像と奥行きマップ	69
5.1	Light field rendering における焦点ぼけ（文献 [14] より引用）	75
5.2	基本設定	77
5.3	“合焦判定”のための二つの合成モード	79
5.4	提案手法のフローチャート	82
5.5	正方格子カメラアレイによる入力画像群（「筑波大学多視点画像データベースより」）	84
5.6	正方格子カメラアレイ画像を用いた実験結果	84
5.7	入力多視点画像群（カーネギーメロン大先端マルチメディア研究室提供）	85
5.8	ある視点位置に対する合成過程：(a)–(c) 平面モデルの奥行きを変化させながら合成した画像，(d)–(f) 参照画像，(g)–(h) 差分画像，(j) 割り当てられた奥行き，および (h) 最終的な“全焦点”画像	86
5.9	他の視点に対する合成結果	87
5.10	ステージごとの処理時間の内訳	87
5.11	従来の image-based modeling and rendering の処理フロー図	89
5.12	視点依存型 image-based modeling and rendering の処理フロー図	89
5.13	本章の提案手法の処理フロー図	89
6.1	多視点画像信号の周波数解析	95
6.2	仮定される奥行きと再構成フィルタ	96

6.3	周波数領域における各フィルタの形状	97
6.4	異なる再構成フィルタによる合成画像	98
6.5	入力多視点画像 (「筑波大学多視点画像データベースより」)	101
6.6	“合焦” 距離が異なる合成画像 (normal filter による)	102
6.7	“合焦” 距離が異なる合成画像 (wide-aperture filter による)	102
6.8	“合焦” 距離が異なる合成画像 (camera-skipped filter による)	102
6.9	合焦判定による奥行き推定	103
6.10	推定された距離画像	103
6.11	提案手法による“全焦点” 画像	103
6.12	他の視点に対する“全焦点” 画像	103
6.13	CG シーンを用いた合成 ($N = 11$)	104
6.14	平均値処理のブロックサイズ M と PSNR の関係	105
6.15	係数 $a_{i,j}$ の組合わせと合成品質	106
6.16	合焦距離の候補数 N と PSNR の関係	106
6.17	周波数領域におけるフィルタ	108
6.18	U と $s-u$ の関係	108
6.19	空間領域における各フィルタの実装	110
7.1	LIFLET の撮像系	113
7.2	撮影された IP 画像の例 (上) と中央付近の拡大図 (下)	114
7.3	LIFLET 光学系を用いた light field rendering	116
7.4	LIFLET 光学系における“合焦判定”	117
7.5	レンズアレイを用いた撮像系 (上) と focal plane (z_1-z_8) の設定の最適化 (下)	118
7.6	画像合成の過程	120
7.7	実験風景	121
7.8	マルチスレッドを用いた実装	122
7.9	処理速度	122
7.10	動空間に対象にした合成の様子	123
7.11	提案手法 (左) と従来手法 (右) の画質の比較	123
7.12	本研究室において開発しているカメラアレイ	124
8.1	light field の周波数解析	130
8.2	仮定される奥行きとフィルタの形状	131
8.3	サブサンプリングされた light field の周波数特性 ($N = 4$)	132
8.4	再構成フィルタ: (a) 焦点が合う場合, (b) 焦点から外れる場合	134
8.5	画像合成の実装	136
8.6	“合焦” 判定法の処理フロー図	139
8.7	視点依存型 image-based modeling and rendering の処理フロー図	139

表目次

2.1	Plenoptic function の縮退とイメージベースレンダリング (文献 [18] より転用)	10
3.1	各理論の比較：被写界深度との関係が図示されているパラメータ	45
3.2	各理論の比較：空間周波数の要素	45
3.3	実験条件	47
4.1	視点依存形状推定手法の比較	58
4.2	画像合成に要する総処理時間 (ms): 81 枚の入力画像を使用	68
4.3	各レイヤに対する処理時間の内訳 (ms)	68
4.4	画像合成に要する総処理時間 (ms): 25 枚の入力画像を使用	68
6.1	CG を用いた実験の諸条件	104
6.2	係数 $a_{k,l}$ の組み合わせ	105
7.1	各種パラメータ	118

第1章

序論

1.1 本研究の目的

近年、視覚情報メディア技術は長足の進歩を遂げている。その進歩に伴い、従来の平面的な二次元画像だけではなく、我々が生活する三次元実環境そのものを扱うことも夢ではなくなった。カメラデバイスの低廉化に加え、様々な撮影装置や撮像方式が開発され、三次元空間の視覚情報を取得するための環境が整いつつある。コンピュータビジョン技術は、三次元環境の構造を理解するための方法論を提供し、ますますその完成度を高めている。コンピュータグラフィックス技術は、計算機内部に構築された精巧な仮想三次元世界を鮮明かつ高速に描き出すことを可能にした。三次元ディスプレイ技術は、観察者に立体感を与えるのみならず、視点位置の移動に伴って連続的に三次元空間を提示することも可能にした。これらの三次元視覚情報処理技術は、従来の二次元画像では実現できなかった、より高度な臨場感を表現するポテンシャルを持っている。近い将来には、バーチャルリアリティ、遠隔地通信、映像コンテンツ制作などの多様な分野で積極的に応用されることが期待される。

本論文では、実空間の三次元視覚情報を表現する一つの手段として、実写多視点画像を用いた自由視点画像合成技術に着目する。一般に、この技術は、分析と合成という二つの段階によって構成される。分析の段階では、実写多視点画像を分析してシーンの構造（幾何形状など）が推定される。これは、コンピュータビジョン分野におけるイメージベーストモデリングの問題である。一般に、この段階は膨大な計算を必要とするため、オフラインで処理されることが多い。合成の段階では、与えられた視点位置からのシーンの見え方を計算して画像を合成する。これは、コンピュータグラフィックス分野におけるイメージベーストレンダリングの問題である。この段階では、分析の段階で得られた情報を、事前の知識として利用する。本論文では、全プロセス（多視点画像の取得から自由視点画像合成まで）を実時間で行うことを想定し、全体を最適化する目的から、上記の二段階を統合的に捉える視点に立つ。本論文の目標は、カメラアレイによる多眼画像を入力として用いて、“全焦点”自由視点画像を実時間で合成することである。なお、本論文では、入力画像には、物理カメラの特性に基づく焦点ぼけがないものと仮定し、“焦点”や“合焦”という言葉を用いて、合成画像上に現れる現象を指して用いる点に注意されたい。

本論文の主たる成果の一つは、新しい画像合成手法の提案と実践である。本論文では、二つの合成手法 – 多眼ステレオ法を最適化した合成手法と、独自の方法論である“合焦”判定に基づく合成手法 – を提案する。“合焦”判定に基づく手法は、シーンが一枚の平面で近似できると仮定して自由視点画像合成を行う手法である、light field rendering 法を拡張したものである。この提案手法では、まず仮定平面の位置を連続的に移動しながら複数の画像を合成する。これらの画像では、仮定平面の位置に応じて、“焦点”ぼけのような現象が現れ、非“合焦”領域は不鮮明になる。次に、それらの画像の“合焦”領域を選択（“合焦”判定）して、一枚の画像に統合することで、全体が鮮明な“全焦点”画像を得る。二つの提案手法はいずれも、グラフィックスハードウェアの機能を用いて効率的に実装され、蓄積された多眼画像データを用いた実時間動作が確認できた。後者の手法は、レンズアレイを用いた実システムに適用され、動空間を対象とした実時間処理（多視点画像の撮像から画像合成まで）が可能であることが確認された。

本論文のもう一つの重要な成果は、理論的体系化である。本論文では、まず、最も単純な自由視点画像法を想定した上で、合成品質の限界を直感的に表現する“被写界深度の理論”を導入する。次に、実時間の自由視点画像合成という目的のもとで、多眼ステレオ法を最適化する方法を検討する。この検討は、合成画像の座標空間で直接問題を解くことで処理の効率化を目指すものであり、一つ目の提案手法に応用される。もう一つの提案手法においては、奥行き推定の新しい方法論として、“合焦”判定法を導入する。“合焦”判定法を空間領域で解析することにより、任意のカメラアレイ入力に適用可能な実用的なアルゴリズムが導かれる。周波数領域の解析により、“合焦”判定法に対して、多次元信号処理の理論に基づくユニークな理論的裏づけが与えられる。“合焦”判定法の意義は、自由視点画像合成のための奥行き推定という問題に対して、入力画像の代わりに合成画像を分析する、という新しいアプローチを試みる点にある。さらに、“合焦”判定法によって導入される理論体系を一般化することで、“合焦”判定法と従来の多眼ステレオ法の両者を共通の枠組みの中に位置づけることが可能である。

本論文で提案する画像合成手法は、実時間で高品質な自由視点画像合成を実現するものであり、今後、合成品質や処理速度を高めるためにより洗練され、遠隔地通信など動空間を対象とするリアルタイムアプリケーションに応用されることが期待される。また、本論文の理論的検討は、自由視点画像合成という問題に対して、コンピュータグラフィックス、コンピュータビジョン、および信号処理という三つの学問分野の境界を超えたアプローチを試みるものである。このアプローチは、関連分野の研究が相互に比較され、関連付けられ、統合される、三次元視覚情報処理という新たな研究分野の確立への道筋を示すものとなることが期待される。

1.2 本論文の構成

本論文の構成は次の通りである。

第1章 序論

第2章 研究の背景

第3章 自由視点画像合成におけるサンプリングと被写界深度の理論

第4章 全焦点自由視点画像合成のための多眼ステレオ法の最適化

第5章 合焦判定に基づく全焦点自由視点画像合成 – 空間領域における検討

第6章 合焦判定に基づく全焦点自由視点画像合成 – 周波数領域における検討

第7章 合焦判定に基づく全焦点自由視点画像合成法の実システムへの適用

第8章 多次元信号処理から見た合焦判定法と多眼ステレオ法

第9章 結論

図 1.1 は、各章の関係を示したものである。以下、本論文の流れを簡単に説明する。

第2章では、本研究の背景について述べる。本論文の主題である、多視点画像を用いた自由視点画像合成について、以下の四つの観点から関連研究を紹介する。1) 画像通信の高度化、2) コン

多眼画像を用いた全焦点自由視点画像の実時間合成に関する研究

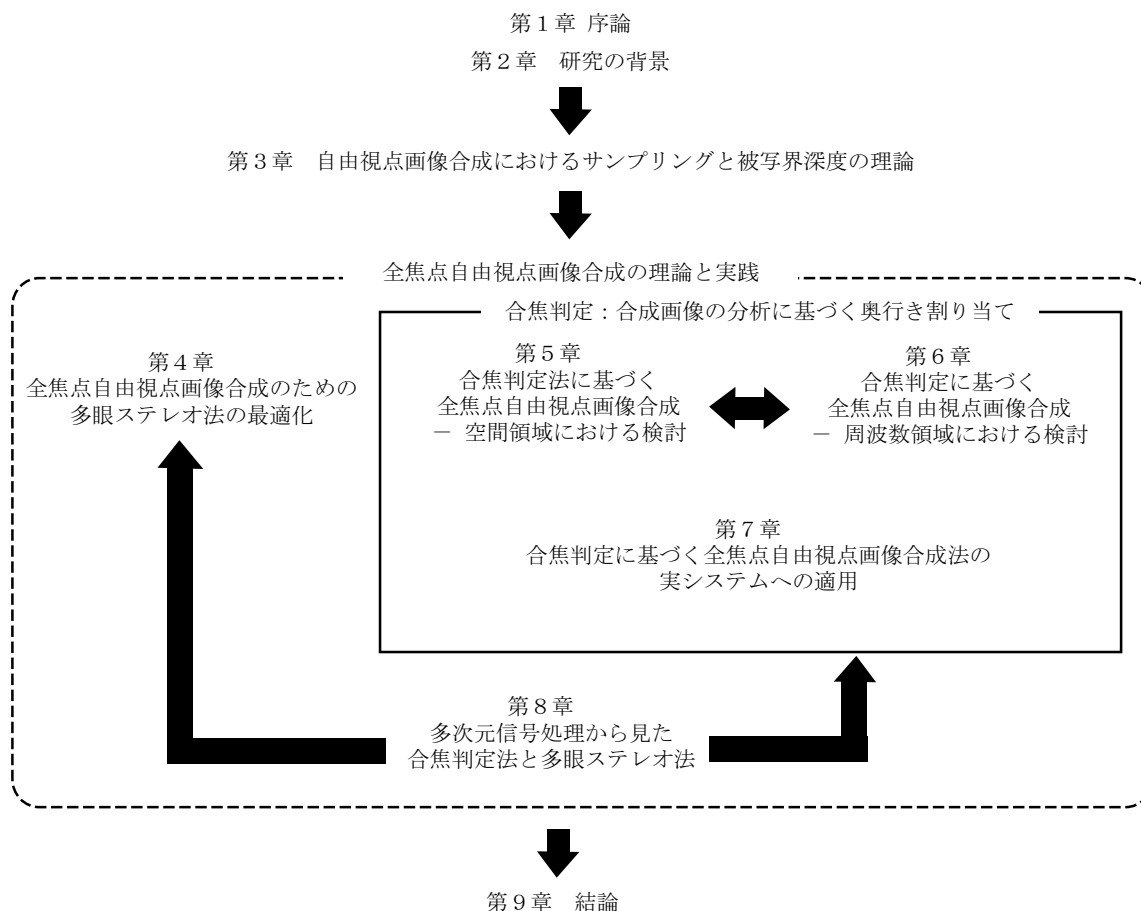


図 1.1: 本論文の構成

ピュータグラフィックスにおけるイメージベースレンダリング, 3) 動空間を対象とした自由視点画像合成システム, および 4) 多次元信号処理.

第3章では, 多眼画像を用いる自由視点画像合成において, 被写体空間を一枚の平面で近似する最も基本的なケースを想定し, 鮮明に合成できる奥行き範囲, すなわち“被写界深度”について考察する. 特に, “被写界深度”とサンプリング条件(多眼カメラのカメラ間隔や画素間隔)の間に成り立つ定量的関係の導出を試みる. この議論は, 自由視点画像合成における平面モデルの限界を示すものであり, 本論文で目指す, “全焦点”合成のための理論的基盤となる.

第4–8章では, 被写体空間を複数の奥行きレイヤ群で表現することで, “全焦点”自由視点画像を合成することを目指し, 理論と実践を展開する. レイヤ群を用いる場合には, 合成画像の各画素を最適なレイヤに割り当てる必要があり, この処理をいかに行うかが本論文の主要な論点となる. これは一種の奥行き推定の問題と考えることができる.

第4章では, 実時間で“全焦点”自由視点画像を合成する, という目的に対して, 奥行き推定の手法である多眼ステレオ法を最適化する. 具体的には, 処理を所望の視点位置に対する合成に

必要十分な範囲に限定することで、コストの削減を図る。また、自由視点画像合成のための奥行き推定に求められる四つの基本的な条件を挙げ、それらを満たす奥行き推定手法の提案と実装を行う。

第5章では、合成画像の各画素に対して最適な奥行きレイヤを割り当てるための新たな方法論を導入する。これは、ある奥行きに設定された単一の平面モデルに基づいて合成された画像を分析することで、各画素に対するその奥行きの妥当性を評価する手法である。合成画像上で鮮明な部分を判定する処理であることから、本論文ではこれを“合焦判定”と呼ぶ。これは、「合成に基づく分析」という新しいアプローチであり、奥行き推定手法の枠組みを大きく広げるものである。また、この手法を用いた自由視点画像合成手法を実装し、有効性を確かめる。尚、本章では、すべての議論は空間領域で展開される。

第6章では、第5章で導入した“合焦”判定法について周波数領域で議論する。本章の議論は、理想的な多眼カメラアレイ入力に限定されるが、第5章と等価な結論を導くものである。本章の議論は、奥行き推定の問題に多次元信号処理の理論（plenoptic sampling 理論）を適用するユニークな試みである。また、CG シーンを用いた実験により、合成画像の品質を定量的に評価し、提案手法の有効性を確かめる。

第7章では、第5, 6章で検討した“合焦”判定に基づく“全焦点”自由視点画像合成法を、レンズアレイを用いた実時間システム LIFLET (Light Field Live with Thousands of Lenslets) に適用する。LIFLET では、特殊な撮像光学系が用いられているため、合成手法やパラメータをシステムに合わせてチューニングする。本章の検討を通して、提案する画像合成手法により、動空間を対象とした実時間合成が可能であることを確かめる。

第8章では、多次元信号処理の観点から、自由視点画像合成のための奥行き推定に関する理論的体系化を行う。第4章で検討した多眼ステレオ法は入力多眼画像を分析する手法であり、第5-7章で検討した“合焦”判定法は合成画像を分析する手法であった。本章の議論では、多眼ステレオ法が、“合焦”判定法の特殊なケースとして、共通の枠組みの中に位置づけられることを示す。

第9章では、本論文の成果をまとめるとともに、今後の課題と展望に触れ、全体の結論とする。

第2章

研究の背景

2.1 はじめに

本論文では、カメラアレイによる多眼画像を入力とし、実時間で、高品質な自由視点画像を合成することを目標とする。多眼画像の取得から自由視点画像合成までを実時間処理できれば、動空間を対象とするリアルタイムアプリケーションへの応用が可能になる。したがって、本論文の目標は、高度な画像通信を実現するための要素技術として捉えることもできる。さらに、本論文における検討は、コンピュータグラフィックスの分野におけるイメージベースレンダリングの問題であり、コンピュータビジョンにおけるイメージベースモデリングの問題であり、かつ多次元信号処理の問題としても捉えられる。

本章では、このような研究の背景について、下記の四つの観点からまとめる。まず、2.2 では、画像通信の高度化という観点から、三次元視覚情報へと関心が高まっていった経緯を述べる。2.3 では、コンピュータグラフィックスの方法論という観点から、実写画像に基づく自由視点画像合成技術であるイメージベースレンダリングについて概説する。2.4 では、システム的な観点から動空間を対象とした自由視点画像合成システムの実例を紹介する。2.5 では、自由視点画像合成の問題を信号処理の立場から解析した既存研究を紹介する。2.6 で本章を締めくくる。

2.2 三次元統合画像通信と光線空間法

次世代の視覚メディアにおいては、平面的な二次元画像のみならず、三次元空間そのものを対象とする、より高い臨場感を持つ視覚情報表現が期待される。したがって、「画像情報のみならず、空間情報の伝送が可能な柔軟性の高い統合的なシステム」[1]が必要とされている。本節では、画像通信の高度化という観点から、三次元視覚情報の表現法として「光線空間法」が生み出され、さらにそれが自由視点画像合成という用途へと応用されてきた経緯を概観する。

人に立体感を与える画像提示に関する試みは数百年の歴史を持つ。特に19世紀から20世紀前半にかけて、写真技術の発達を背景に、両眼視差を与えるための技術が開発されてきた。立体鏡、アナグリフ方式、偏光方式などのメガネを用いるものから、レンチキュラスクリーン方式、パララックスバリア方式、インテグラルフォトグラフィ方式のようにメガネ不要のものまで、様々な方式が次々に提案されている（立体表示の諸方式については、文献[2, 3]などを参照されたい）。しかし当時の周辺技術の未熟さから、これら優れた着想の中には実現されなかったものもある。1990年代になって、2次元画像表示の高画質化を突き詰めたHDTVの登場を受け、三次元映像表示技術が次世代を担う技術として注目を集めた。このころには、とりわけデジタル信号処理技術が新たな道具として加わったことにより、100年前には写真技術でしかなかった三次元表示技術は、さまざまな形式の三次元ディスプレイとして実用化されるようになった[4, 5, 6, 7, 8]。

上記のような進展を背景に、原島らは「三次元次元統合画像通信」なる枠組みを提唱した[9, 10, 11, 12]。三次元画像通信の目的は、三次元空間の情報を取り出して伝送し、受信側の観察者にあたかもその空間にいるかのような臨場感を与えることにある。しかし、当時の三次元ディスプレイの研究は、ハードウェア的な画像表示技術や奥行き知覚などの視覚特性に関するものが大部分であった。三次元情報を取得する方式（入力方式）および、表示する方式（出力方式）は多種多

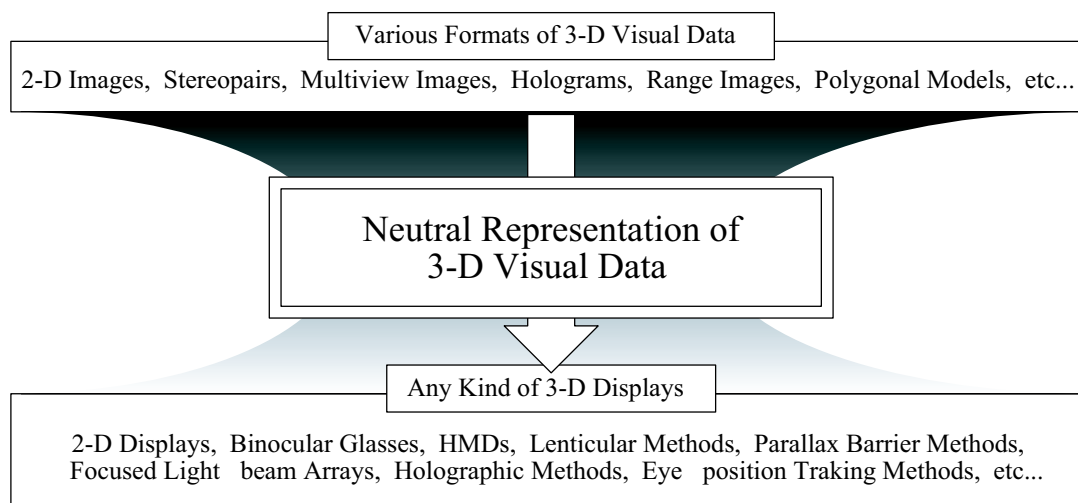


図 2.1: 三次元統合画像通信の概念図（文献 [1] より引用）。

様であり，相互の互換性についてはほとんど検討されていなかった．原島らは，図 2.1 に示されるような，入出力デバイスに依存しない中立的なデータフォーマットを導入することで，これらを統合的に扱う枠組みを与えることができると考えた．この構想の下に，三次元視覚情報の最小単位として「光線」が注目され，光線をパラメータ化した情報空間である「光線空間」という考え方が生まれた [13, 14, 15]．光線空間法では，空間を満たす光線群は，光線の通過位置 (x, y, z) と方向 (θ, ϕ) をパラメータとして，5 次元の関数 $f(x, y, z, \theta, \phi)$ によってことごとく記述される．この記述方式は中立的なものであるため，どんな入力デバイスで取得された情報も，光線空間のフォーマットに変換でき，またこのフォーマットから任意の出力形式へと変換できる．

このように 3 次元空間を光線の情報で表現した場合，画像の撮影とは「光線情報のサンプリング」を行う操作である．一方，「ある視点を通る光線情報の取り出し」を行って画像として表示する操作は，自由視点視点画像の合成に相当する（「ある視点」は任意に制御できる）．こうして当初は三次元ディスプレイを想定していた多眼画像処理・符号化技術は，自由視点画像の合成という新しい境地に到達した．特に，1990 年代後半以降には，コンピュータグラフィックスの分野で，実写画像を用いた写実的な画像合成法である，イメージベースレンダリングが注目されるようになり，多くの研究者がこの分野に関心を持つようになった．「光線」を単位として三次元視覚情報を扱うという点で，光線空間法とイメージベースレンダリングは本質的に共通の基盤を持つものである．したがって，「三次元統合画像通信」および「光線空間法」の基本的な考え方は，形を変えながら，その後の研究の中に脈々と生き続けているといえよう．

2.3 イメージベースレンダリングについての概論

イメージベースレンダリングとは，主にコンピュータグラフィックスの分野で用いられる用語であり，あらかじめ取得された多数の画像から，新たに自由視点画像を合成するさまざまな手

法を現す．幾何学モデルに基づく従来の伝統的なグラフィックス手法と比較して，より少ない計算量で，高い写実性を実現できるため，近年多くの研究者の注目を集めている．これらの手法の中には，画像群のみを入力データとし，幾何情報を全く用いない手法もあるが，大半は，幾何学的な情報を相補的に用いることにより，データ表現の効率性と品質の高い画像合成を目指すものである．

本節では，文献 [16] に倣い，イメージベースレンダリングの多様な手法を，幾何情報への依存度という観点で下記の三つのグループに分類し，代表的な手法について概説する（2.3.1–2.3.3）．

1. 幾何情報を用いない手法 (rendering with no geometry)
2. 暗示的に幾何情報を用いる手法 (rendering with implicit geometry)
3. 明示的な幾何情報を用いる手法 (rendering with explicit geometry)

ただし，この分類は，説明の都合上の便宜的なものであり，グループ間の厳密な線引きは極めて困難である点を書き添える．さらに，2.3.4 では，これらを俯瞰する視点から，必要とされる画像情報と幾何情報のトレードオフ関係について論じる．

2.3.1 幾何情報を用いない手法

まず，イメージベースレンダリングの最初のグループとして，幾何情報を用いない手法について述べる．これらの手法では，被写体空間の「見え」を被写体空間から飛来する光線の集合として表す．したがって，上記の三つのグループの中では，最も多くの入力画像を必要とする．

「視覚情報は空間を満たす光線を単位として表現できる」というアイデアの起源は，古くはルネッサンス時代にまでさかのぼることができる．レオナルド・ダ・ビンチは，空間のある一点を通過する光線の集合を，“radiant pyramid”という言葉で表現しており，空間は無数の radiant pyramid で満たされている，と述べている．Adelson ら [17] は上記のアイデアに基づいて，空間中の光線を七次元で表現する関数，plenoptic function¹ を導入した．七つの変数は，視点の座標 (x, y, z) ，光線の方向 (θ_x, θ_y) ，波長 λ および時間 t である．この七次元のデータを用いることで空間中のすべての光線を記述することが可能である．ある時間において，ある任意視点からある方向を向いた画像を得たい場合，その方向に画像平面を想定し，その平面を横切るすべての光線を読み出してプロットすることによって所望の画像を得ることができる．

しかし，七次元の plenoptic function は膨大な情報を持ち，扱いが容易ではない．そこで，様々な仮定の元に plenoptic function の次元を減らすことによって，より実用的なイメージベースレンダリング手法が生み出されてきた．Zhang と Chen [18] の分類によると，次元を減らすための仮定は六つある．表 2.1 に示されるように，これらの仮定の組み合わせによって，通常の二次元画像を含め，様々なイメージベースのデータ表現手法を位置づけることができる．

仮定 1 光線の色は三つのチャンネル (R,G,B) によって表現される．波長成分 λ を省略できる．

仮定 2 光線は直進しその方向に沿って減衰や干渉がなく色が変化しない．次元が一つ減る．

¹plenoptic という言葉は，「完全な，すべての」を意味するラテン語の plenus と，見えに関係する optic から来ている．

表 2.1: Plenoptic function の縮退とイメージベースレンダリング (文献 [18] より転用) .

Dimension	Example	Assumptions
7D	Plenoptic function	–
6D	Surface plenoptic function	(2)
5D	Plenoptic modeling	(1,3)
	Light field video	(1,2)
4D	Light field/lumigraph	(1,2,3)
3D	Concentric Mosaics	(1,2,3,4)
	Panoramic video	(1,6) or (1,3,5)
	Branch movies	(1,3,5)
	Video	(1,6)
2D	Image mosaicing	(1,3,6)
	Image	(1,3,6)

仮定 3 静的なシーンである．時間軸 t が省略できる．

仮定 4 視点位置の動きが二次元平面上に限定される．次元が一つ減る．

仮定 5 視点位置の動きが特定の線上に限定される．次元が二つ減る．

仮定 6 視点位置が動かない．次元が三つ減る．

仮定 2 では，光線の色が変化しないことを条件とした．この条件下では，空間全体を，被写体の存在する空間（記述する空間）と観察者の視点がある空間（記述される空間）に分けるような境界面を考え，その面を通過する光線のみを記録すればよい．したがって，光線の位置は三次元ではなく二次元で表現でき，結果として次元を一つ減らすことができる．

表 2.1 において，高次元の手法から順に説明してゆく．まず，Surface plenoptic function[19] は，物体の表面上で定義された plenoptic function である．仮定 2 により，六次元の表現となっている．すなわち，表面上の位置（2D），光線の方角（2D），時間（1D）および波長（1D）である．Surface plenoptic function は，光線情報の理論的分析において便利なツールではあるが，概念的な表現であり，レンダリング手法としての実用には至っていない．

波長成分と時間成分を省略することにより，McMillan と Bishop [20] は plenoptic modeling を導入した．データ形式は，三次元空間中の各点（3D）で，円柱状の全方位画像（2D）が与えられている，というものである．彼らは，撮影位置を三次元的に変えつつ，カメラをパン（水平方向の回転）させながら繰り返し画像を撮影することにより，その五次元空間を満たすデータを取得した．また，与えられた任意の視点位置に対して，近接する円柱画像を，エピポーラ幾何関係に基づいてワーピングすることにより，自由視点画像合成を実現した．異なる仮定の組み合わせを用いて五次元の情報空間を扱うものとして，light field video [21, 22] がある．これは，後述する light field に時間軸を加えたものである．

四次元の形式としてよく知られているのは，light field[23] と lumigraph [24] である．これらの

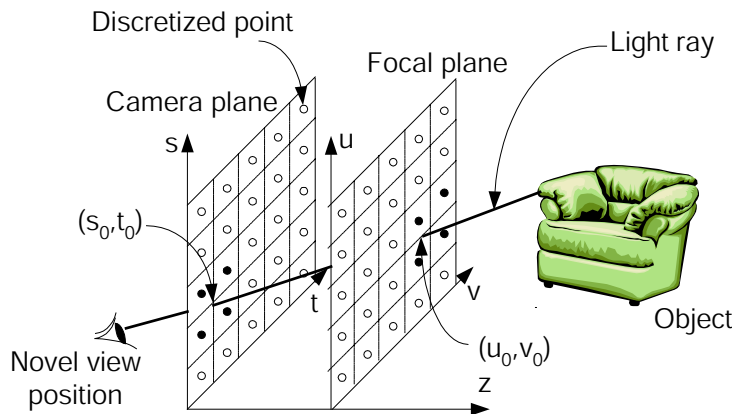


図 2.2: Light field のデータ形式とレンダリング (文献 [18] より引用) .

手法では，光線の波長と時間成分を無視し，さらに光線の色が変化しないという仮定を置いている．具体的なデータ形式は，図 2.2 に示すように，二枚の平面を定義し光線をそれらの平面との交点で表現するというものである．すなわち一方の平面との交点を (s, t) ，他方との交点を (u, v) とすることで，光線は四つのパラメータの組 (s, t, u, v) によって一意に識別される．このデータ形式は実用的である．例えば，カメラアレイによって取得された多視点画像は，離散化された light field データとみなすことができる．任意の視点位置 (Novel view position) で画像を合成するためには，図 2.2 のように，その視点位置に到達する各光線の色を求めればよい．このとき，サンプリング点の間を通過する光線については，近傍のデータを使って適切に補間する．前述の光線空間法 [15] も，具体的なデータ表現の違いはあるものの，原理的に等価な手法とみなせる．

上記の四次元のデータ形式では，画像を合成する視点位置を三次元的に動かすことができる．視点位置の移動範囲を二次元面上に限定することにより，次元を一つ減らした手法として，concentric mosaics [25] がある．この手法では，図 2.3 に示すように，カメラを円状に回転させながら画像を取得し，角度 (1D) と画素位置 (2D) によってデータにパラメータづけする．得られたデータは，一列に並べたスリットカメラを同心円状に回転させながら撮像することによって得られるモザイク画像群と等価であり，concentric mosaics と呼ばれる．画像合成は，図中の rendering circle 内部の任意の視点位置で可能である．このとき，合成画像の縦方向の歪を補正するため，被写体の奥行きを仮定した処理を導入している．

自由視点画像合成という枠組みを外れるが，三次元のデータ形式を用いるものとしては，branch movies [26, 27, 28]，パノラマビデオ (panoramic video) [29, 30, 31]，および通常のビデオ，二次元のデータ表現を用いるものとしては，モザイクキング画像 (image mosaicing) や通常の画像を挙げることができる．Plenoptic function の次元縮退という観点からは，これらの方式はイメージベースレンダリングと共通の枠組みに位置づけることが可能である．特に，パノラマビデオやモザイクキング画像については，視点位置は固定されるものの，複数の実写 (動) 画像を「継ぎ合わせる」処理が行われるため，広い意味ではイメージベースレンダリングに含めることが多い．

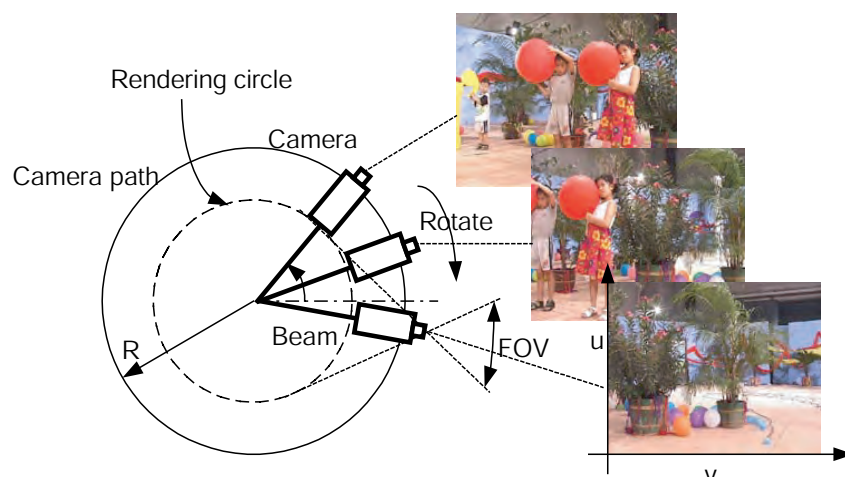


図 2.3: Concentric mosaics (文献 [18] より引用) .

2.3.2 暗示的に幾何情報を用いる手法

ここで述べるのは、比較的少数の入力画像とそれらの画素位置の対応関係を用いる手法である。これらの手法は、「暗示的」に幾何情報を用いている。なぜなら、この画素位置の対応関係のみから対象の幾何情報を直接得ることはできないからである。三次元形状の推定を行うためには、画素位置の対応関係に加えて、各画像のカメラパラメータが必要である。これらの手法は、明示的な幾何情報を用いる代わりに、画素位置の対応関係を直接利用して自由視点画像合成を行う。

View interpolation [32] では、二枚の画像と、その間の画素位置の対応関係 (optical flow) を用いて、自由視点画像を合成する。この手法が十分に機能するためには、二枚の入力画像の撮影地点の距離が近く、かつカメラの向きがほとんど同じである必要がある。撮影地点間の距離が離れていると、オクルージョンの問題が顕著になり、かつ二枚の画像間で共通しない部分が多くなる。また、この手法を実写画像に当てはめる際に問題となるのは、optical flow の構築 (画素の対応関係を求める処理) が容易ではないことである。

View morphing (図 2.4) [33] は view interpolation の特殊な場合に位置づけられる。View interpolation では合成画像が幾何学的に正しい保証はなかった。一方、view morphing では、仮想視点位置の動きを、二台の入力カメラの光学中心を結ぶ直線上に限定することにより、合成画像の幾何学的正当性を保証している。仮想視点が動く直線と二台の入力カメラの光軸が直交する場合には、合成画像は二枚の入力画像の線形結合として与えられる。もし、入力カメラ同士が平行でない場合には、まず pre-warping により入力画像を補正する。また、合成画像に post-warping を施すことにより、仮想視点カメラを任意の方向に向けることができる。

入力画像間の幾何学的な拘束関係に基づいて自由視点画像合成を行う手法は transfer methods と呼ばれる。Laveau と Faugeras [34, 35] は基本行列 [36] (二枚の画像のエピポーラ拘束を表現する行列) を用いた画像合成法を提案した。入力画像間のエピポーラ拘束のみが与えられている (こ

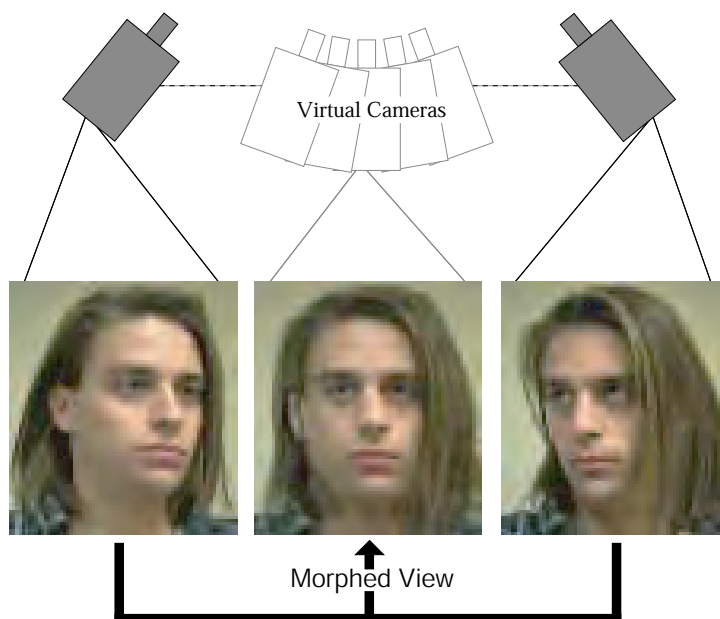


図 2.4: View morphing (文献 [32] より引用)。

の状態を weak calibration と呼ぶ) 場合には, 復元されるのは射影構造 [37] (ユークリッド構造における角度や面積が保たれない) であるため, 合成画像に不自然な歪が出ることが原理的に避けられない。一方で, カメラパラメータがすべて既知 (この状態を full calibration と呼ぶ) の場合, このような歪を回避することができる。また, 三枚の画像間の幾何学的拘束関係である, trifocal tensor [36] を用いた画像合成手法 [38] も提案されている。

2.3.3 明示的な幾何情報を用いる手法

ここでは, イメージベースレンダリングの最後のグループとして, 明示的な幾何情報を用いる手法について述べる。コンピュータグラフィックス技術はもともと対象物の形状情報を出発点にしており, ごく一般的に用いられている, 三次元モデルにテクスチャを貼り付ける手法も, 分類としてはこのグループに属する。

3D warping [39] は, 入力画像の画素を三次元座標に逆投影し, それらを新たな視点位置に再投影することで, 自由視点画像を得る手法である。一枚以上の入力画像で, 各画素の奥行きが既知である必要がある。合成画像上で, オクルージョンの影響によって生じる穴をいかに埋めるか, というのがこの手法における重要な問題である。Oliveira と Bishop [40] は, 比較的単純な pre-warping 処理と, 通常のテクスチャマッピングを組み合わせ, 高速に 3D-warping を実現する手法を開発した。また, Rademacher と Bishop [41] は, 複数個の投影中心を持つ, MCOP (multiple-center-of-projection) 画像に対して, 3D-warping を適用した (図 2.5 参照)。Layered depth images (LDI) [42] は, 3D-warping によって生じる穴の問題を解消するため, シーンを層状のモデルで扱う手法である。この手法では, 入力画像の各画素に奥行き方向に沿って複数の色を持たせることで, 背後に隠れている部分の情報を扱う。さらに, LDI に多重解像度の階層性を

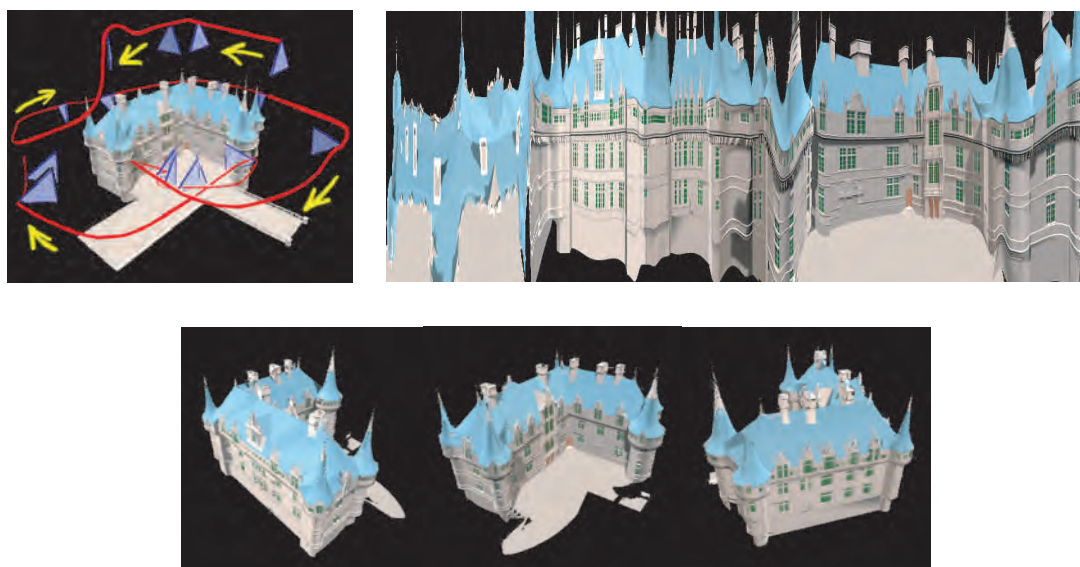


図 2.5: MCOP の撮影（左上），構築（右上）および再投影（下）（文献 [41] より引用）。

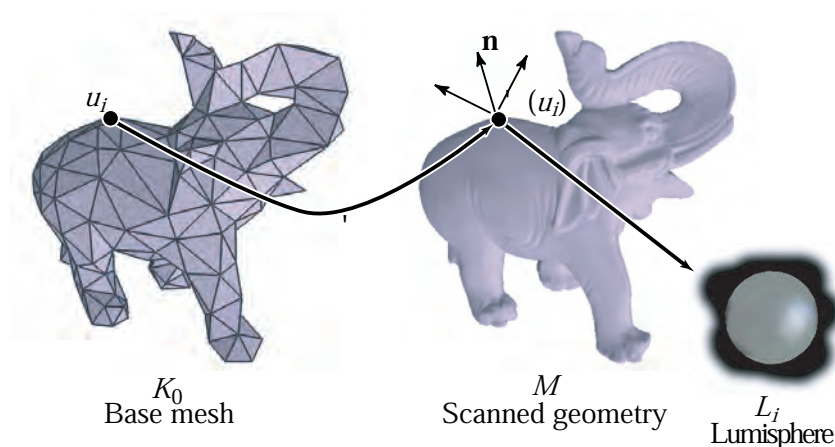


図 2.6: Surface light field（文献 [46] より引用）

持たせた，LDI tree というデータ構造も検討された [43]。

View-dependent texture mapping [44, 45] は，従来のテクスチャマッピングを拡張した概念であり，視線方向に応じた見えの変化（ハイライトや鏡面反射成分）を表現できる．この手法では，ある面に対して，異なる方向から撮影された複数のテクスチャを足し合わせてマッピングするが，観察者の視線方向に応じて，それぞれテクスチャの重みを適宜変化させることで，視点依存の見えの変化が再現される．この考え方の延長上にあるのが，surface light field（図 2.6）[46, 47] である．先に述べた light field [23] では，仮想的な平面を横切る光線がパラメータ付けされたが，surface light field では，物体の表面から各方向に放射される光線がパラメータ付けされる．さらに，eigen texture 法 [48, 49] では，光源の変化を含めた surface light field を効率的にパラメータ

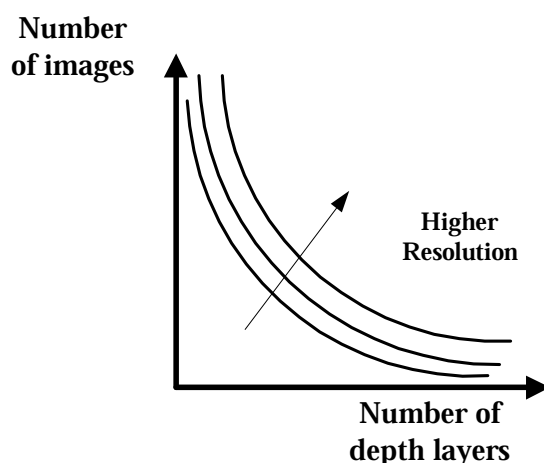


図 2.7: Minimum sampling curve (文献 [53] より引用)

化する手法が提案されている。

さらに比較的新しいアプローチとして、観察者の視点や視線方向に応じて、形状モデル自体を変化させる試みもなされている。Kang ら [50] は、正確な三次元形状が得られない部分については、観察者の視点位置に応じて、複数の視点位置で得られた幾何情報を適応的に用いることが有効であることを示した。Heigl ら [51] は、同様に局所的なデプスマップを用いることが、ハイライトや透明性などの視覚効果を表現する点で有効であることを示した。Micro-facet billboard [52] は、従来のポリゴンベースの表現に向かない材質の物体（毛皮など）の物体や、幾何形状の精度が不十分な物体を表現することを目的とした手法である。この手法では対象物の表面を微小面の集合で表現する。各微小面のサイズは与えられた幾何形状の精度によって決定される。これらの微小面は観察者の視線方向に応じて回転（billboard）され、面と視線が常に直交するように配置される。各微小面に対して light field が定義され、視点に応じた見え方が再現される。

2.3.4 画像情報と幾何情報のトレードオフ関係

本節では、イメージベースレンダリングの手法を大きく三つのグループに分けて、それぞれの代表的な手法を紹介してきた。幾何情報を用いない手法では、膨大な量の画像データを必要とする。一方、幾何情報を用いる手法では、それほど多くの画像を用いなくても、満足な品質を達成できることが多い。一般にある品質を達成するために必要な画像データと幾何データはトレードオフ関係にあることが知られている。Chai ら [53] はこの関係を、minimum sampling curve（最小サンプリング曲線）として定式化している。図 2.7 のグラフにおいて、縦軸は画像の数、横軸は幾何情報の奥行き方向の分解能を表す。合成画像の解像度を一定としたとき、両者に反比例の関係があることがわかる。

ただし、Chai らの理論的解析は、非ランバート反射やオクルージョンの影響を無視するという仮定を置いており、これは幾何情報ベースの手法に有利な仮定であることを付け加えなければならない。Plenoptic function [17] を再現するという観点で考えると、入力画像の取得密度は高い

ほどよく、取得密度が不十分な場合にはデータの補間を行う必要がある。幾何情報を用いるのはより「もっともらしい」補間を行うための手段である。しかし、補間値はあくまでも推定であり、同じ位置で実際に取得される値と同一であることは保証できない。補間処理に過度に依存すると、元の plenoptic function の性質が十分に再現できない。これは特に、視点依存の見え（鏡面反射やハイライト）に顕著に影響する。したがって、より厳密な意味では、幾何情報の精度を上げれば画像の取得密度の不足を補える、という論理は正しくない。

さらに、幾何形状データの取得は容易ではないことも考慮に入れる必要がある。形状データを得る一つの方法は、取得された画像群に各種コンピュータビジョン手法（ステレオ法 [54, 55] やシルエット法 [56] など）を適用することである。しかし、これらの手法で、十分な精度の幾何情報を得ることは極めて困難であり、推定誤差は合成画像の視覚的品質に大きく影響する場合もある。より正確に形状データを得るには、アクティブなレンジセンサ技術（レーザ光やプロジェクタ光を対象に照射する手法）を用いて三次元計測を行う [57, 58, 59] 必要があるが、この計測は手間と時間を要するうえ、計測対象の大きさ、形状の複雑さ、および材質の面で制限が多い。この問題を解決するため、人手による処理を相補的に用いることもある [60]。

2.4 動空間を対象とした自由視点画像合成システムの例

前節では、主にレンダリング手法という観点から、実写画像に基づく自由視点画像合成技術手法全般を概観した。極論すれば、被写体のモデル（見えや形状）を正確に取得し、それをグラフィックスハードウェアに適したデータ形式に変換できれば、自由視点画像合成の問題は解けたことになる。しかし、三次元環境のデータ取得から、分析、データ変換を経て、自由視点画像合成までを一連の流れとして見た場合、各段階に対する個別の検討のみでは研究として不十分である。

本節では、このようなシステムの観点から自由視点画像合成を実現している研究例を紹介する。ただし、本論文の目的に照らして、動空間を対象とし、かつカメラ画像のみを入力とするシステムに絞って概説する。また、説明の便宜のため、オブジェクトを対象とするシステムと、シーンを対象とするシステムに分けて説明する。前者では、背景差分などによって対象となるオブジェクトを切り出すため、背景領域に特殊なマーカー等を設置しておくことも可能である。一方、後者では、画面に映っているシーン全体が処理の対象となる。本論文において主に対象とするのは、後者のシーン全体を処理の対象とするケースである。

2.4.1 オブジェクトを対象とするシステム

三次元動物体を対象にした初期の試みとして最も有名なものは、金出らの virtualized reality システム [61, 62, 63] である。彼らは、被写体の動画像をあらゆる方向から撮影するため、図 2.8 に示されるような三次元ドームと呼ばれる撮影スタジオを構築した。撮影された画像群に多眼ステレオ法を適用することにより被写体の三次元モデルを構築し、そのモデルに色をつけることにより自由視点画像を合成した。ただし、三次元モデルの構築には膨大な計算が必要なため、多視点ビデオの取得と自由視点画像合成の間に、いったんオフライン計算の段階を挟む。このプロジェ

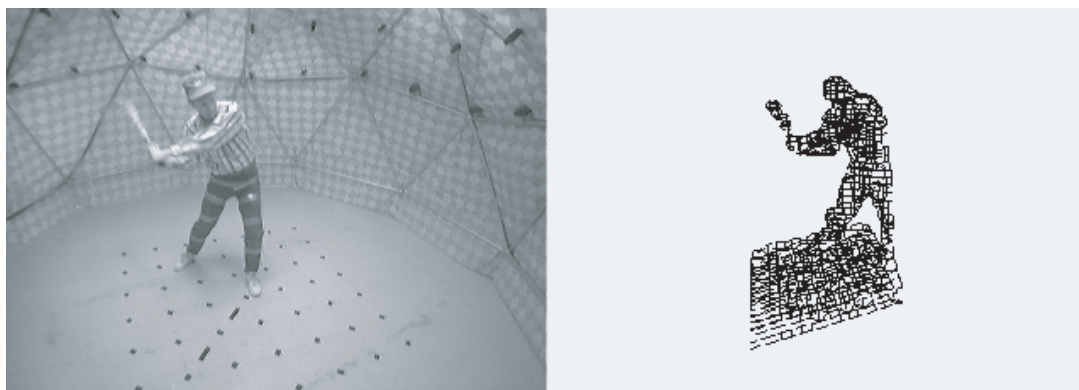


図 2.8: Virtualized Reality システム (文献 [61] より引用) .

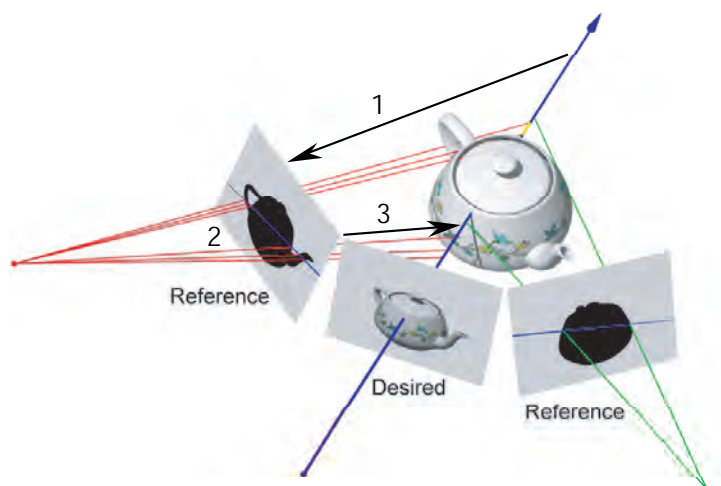


図 2.9: Image-based visual hulls (文献 [68] より引用) .

クトで開発された eye vision システムは，フットボールや野球の中継で使用され注目を集めた．オフライン計算による三次元モデル構築を前提とする方式は，稲本と斎藤 [64, 65] のサッカースタジアムを対象とした自由視点鑑賞システムや，松山ら [66, 67] の三次元ビデオプロジェクトにおいて採用され発展を続けている．

上記のようにオフライン計算を含む方式では，蓄積型のコンテンツを作成することはできても，リアルタイムアプリケーションは不可能である．この問題を解消するため，Matusik ら [68] は，モデリングとレンダリングの両方を実時間で行う，image-based visual hull と呼ばれる手法を提案した．この手法では，シルエットを用いた視体積交差の原理 [56] とエピポラ幾何学を用いて，所望の視点位置から見た形状と色を効率的に計算する．具体的には，図 2.9 に示すように，1) まず所望の視点位置から延ばした光線を参照画像に投影し，2) 次に投影された光線と参照画像のシルエットとの交差部分を求め，3) 最後に交差部分を元の光線に逆投影することで，所望の画像の

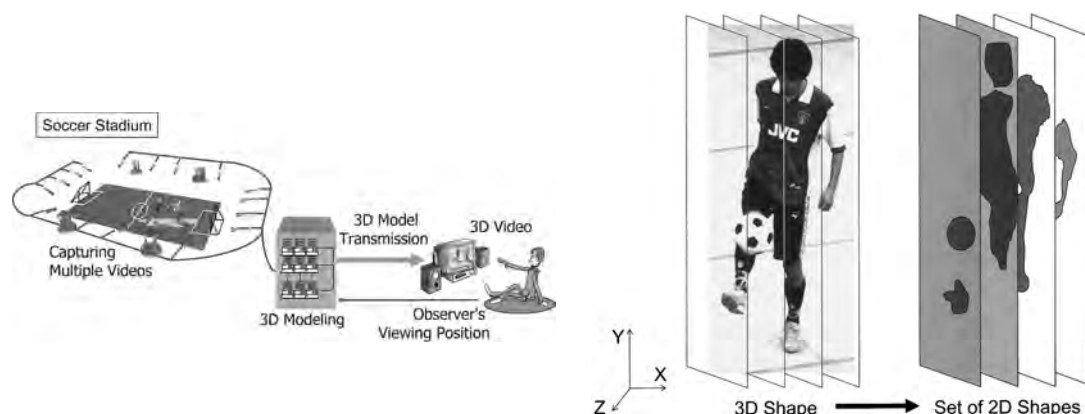


図 2.10: 大規模空間を対象とした三次元ビデオシステムの概念図 (左) と多層平面に基づくスケーラブルな三次元モデル (右) (文献 [76] より引用)。

各画素の奥行きを決めることができる。この手法で推定される幾何情報は、合成対象となる視点位置に依存する (view dependent である)。彼らは 4 台のカメラから成る比較的小規模なシステムを構成し、リアルタイムの動作を確認した。

Image-based visual hull [68] では、形状推定のためにシルエット情報のみを用いているため精度に限界があり、例えば、凹んだ構造を持つ形状は正しく推定できない。形状推定を向上させるためには、色の情報 (photo-consistency) を併せて用いることが有効である。Slabaugh ら [69, 70] の image-based photo hull は、image-based visual hull のアルゴリズムと、画像の色情報を用いる photo hull 復元の手法 [71, 72] を融合し、実時間性を保ちながらより高品質な画像合成を実現したものである。Li ら [73] は、この手法をグラフィックスハードウェア上で効率的に実装することにより高速化し、Slabaugh らの実装の約 7 倍の処理速度を実現した。

北原ら [74, 75, 76] は、サッカースタジアムやコンサートホールなどの大規模な空間を対象とした三次元ビデオを提供することを目指とし、要素技術としてスケーラブルなオブジェクト表現法を提案している。この手法では、オブジェクトは図 2.10 に示すような多層平面で表現され、ユーザの与えた視点位置に対して、三次元モデルの構築と画像合成が実時間で行われる。多層平面データの構築は、shape-from-silhouette 法 [56] で行われる。自由視点画像に特化したチューニングとして次のような工夫を行っている。1) 多層平面は、合成画像の画像面と平行に配置される。2) 多層平面の解像度は、合成画像の解像度と同一にする。3) 多層平面の間隔 (枚数) は、画像を合成する視点位置に応じて変える (遠くから見る場合は、平面は少なくてもよい)。これらのチューニングにより、三次元データのサイズおよび処理時間を、所望の視点位置における画像合成に必要な最小限のレベルに抑えることができる。大規模空間を対象としたアプリケーションでは、多くの物体が存在するため、このチューニングの効果は大きい。

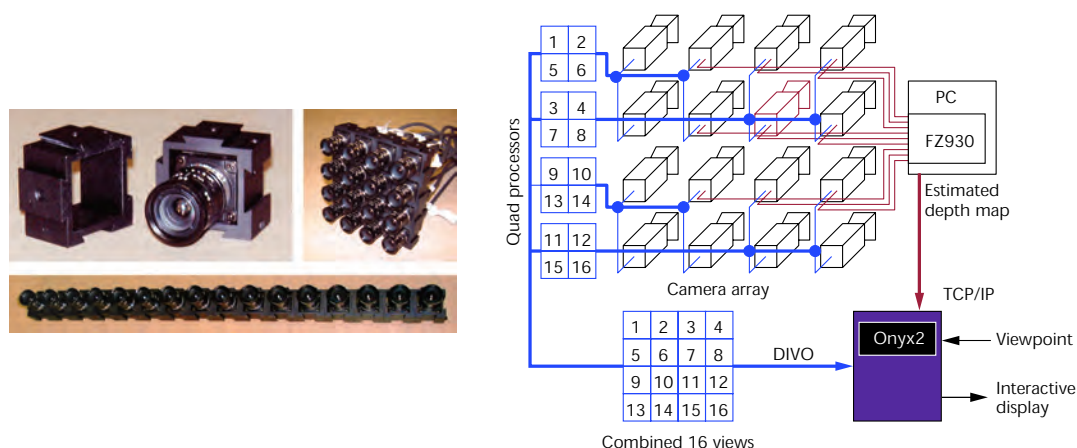


図 2.11: Video based rendering システム (文献 [79] より引用)。

2.4.2 シーンを対象とするシステム

苗村ら [77, 78, 79] は、図 2.11 に示すような 16 台のビデオカメラを用いた実時間自由視点画像合成システムを提案した。静止画を入力とする image based rendering に対して、ビデオ画像を入力とするという意味で、彼らの論文の中では、世界に先駆けて video based rendering という言葉が用いられた。最初のシステムでは、単一の奥行きを仮定して 16 枚のビデオ画像を「継ぎ貼り」するだけの合成であったが、改良システムでは、被写体空間を 2-3 層のレイヤモデルで近似して合成に用いることにより、合成画像の品質を向上させた。改良システムでは、実時間ステレオマッチング機能を持つ専用のハードウェアが用いられており、16 眼のうち 9 眼分の画像を用いて推定された視差マップを量子化することでレイヤモデルが構築される。大井ら [80] は、通常のカメラの代わりにランダムアクセスイメージセンサを用いることにより、このシステムに対する別の改善を試みた。このイメージセンサを用いることで、画素単位でデータの読み出しが可能になるため、読み出しデータを必要十分な量に抑えることができるというメリットがある。

國田ら [81, 82, 83] は、任意視点人物像を合成することを目的とし、12 台のカメラを水平に配置したカメラアレイを構築した。遠隔地通信における応用が想定されたため、当初より、画像の取得から合成までを実時間で行うことが条件とされた。このシステムでは、被写体を単一の平面モデルで近似するため合成品質に限界があるが、形状推定処理は不要であり、画像合成に要するコストも低い。また彼らは、被写体が許容される品質で合成可能となる空間的範囲の制約条件を等価被写界深度として定式化し、システムの設計に盛り込んでいる。さらにカメラアレイを水平方向に往復運動させながら画像を取得することで、時分割処理により仮想的にカメラ数を増やし、等価被写界深度を広げる工夫を行っている。

Yang ら [22] は、動空間の実時間レンダリングを実現するため、図 2.12 に示されるように、64 台 (8×8) の IEEE1394 ビデオカメラでカメラアレイを構築した。國田らと同様、画像合成には単一の平面モデルに基づく手法を用いるため、形状推定処理は不要である。また、大井ら [80] と同様に、画像を合成する視点位置に応じて、各カメラから必要な部分だけを伝送するシステム設計

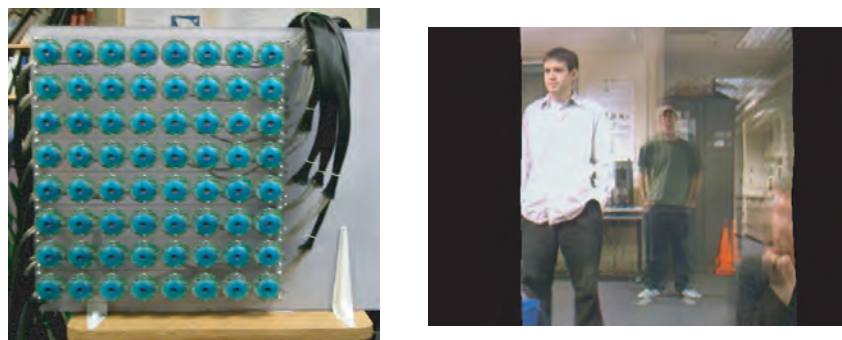


図 2.12: Yang らのカメラアレイシステム (左) と合成画像の例 (右) (文献 [22] より引用)。

を採用した。このとき、必要な通信帯域は、合成画像の解像度とユーザの人数に比例するが、カメラの台数に対しては一定となる。実験では、一人のユーザに対して、80 ms の遅延で 18 fps の速度でのレンダリングが達成された。将来的には、インターネット経由で複数ユーザが同時に利用するようなシナリオが想定されている。

谷本らは、自由視点画像合成技術を次世代のテレビ放送への応用することをめざし、自由視点テレビ (Free Viewpoint Television) プロジェクト [84, 85] を立ち上げ、実時間動作するカメラアレイシステムの試作を行っている [86, 87]。このシステムでは、16 台の CCD カメラ (直線配列、アーク状配列、および二次元アレイ配列が試みられた) で撮影された画像群を元に光線空間 [15] を構成し、適応フィルタ法 [88] (ステレオマッチングによる視差推定と色の補間を組み合わせたもの) によって光線空間を高速に補間し、必要な部分空間を適宜切り出すことにより画像合成を行う。最も計算コストを要する補間処理では、PC クラスタを用いた分散化により高速処理を達成している。また、画像の時系列方向の相関を利用し、変化のない部分については、過去のフレームにおける補間結果を利用することによっても処理効率を高めている。

Zhang と Chen [89, 90] の self-reconfigurable camera array は、シーンの構成や画像を合成する視点位置に応じて自律的に構成を変化させる、能動的な実時間自由視点画像合成システムである (図 2.13 参照)。カメラアレイは 48 台のネットワークカメラと、各カメラを水平方向に平行移動/回転させる駆動装置によって構成され、合成画像の品質を最適化するように各カメラの位置が逐次制御される。各カメラで撮影された画像は JPEG 圧縮され、ネットワークを介して一台の PC に伝送される。PC では、受け取った画像を復号し、レンズ歪を補正した後、メッシュモデルの推定と画像合成を行う。メッシュモデルは、その各頂点に対して space-sweeping 法によるステレオマッチング [71] を適用することにより推定される。さらに、奥行き変化の大きな部分では再帰的にメッシュを分割して新たな頂点を加えてゆき、段階的にメッシュの解像度を向上させる。

山本ら [91, 92] が開発した LIFLET (Light Field Live with Thousands of Lenslets) は、カメラアレイの代わりに、数千個のレンズを配列したレンズアレイを用いた実時間システムである。撮像系の外観を図 2.14 に示す。このシステムの特徴の一つは、カメラアレイを用いる他のシステムと比較して規模やコストが小さいことである。レンズアレイを通して撮影された多視点画像は、カメラアレイを用いて取得された多視点画像と本質的に等価である。このシステムでは、文献 [89]

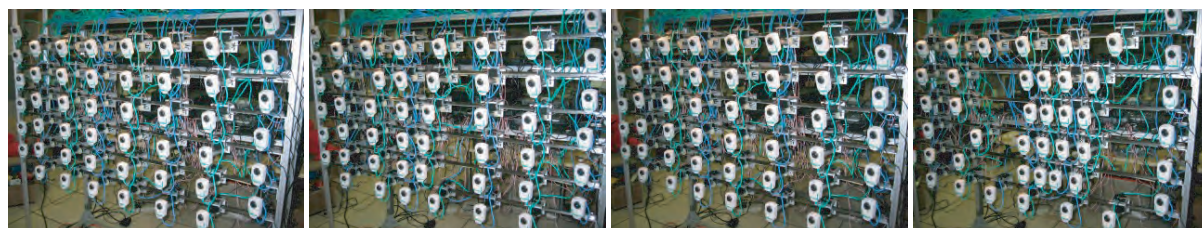


図 2.13: Self-reconfigurable camera array (文献 [89] より引用)。

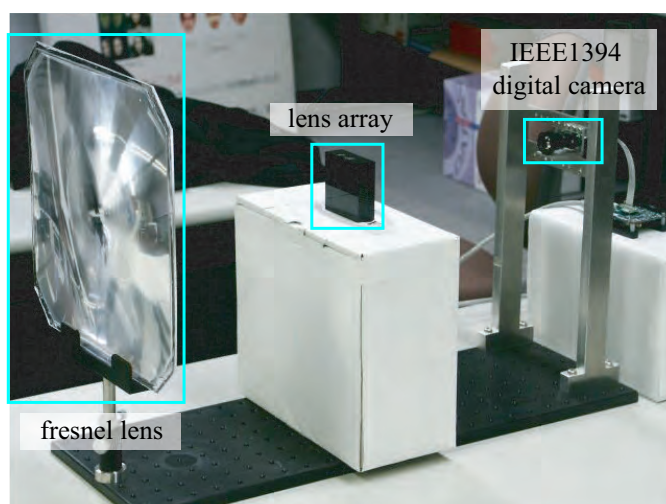


図 2.14: LIFLET システム (文献 [91] より引用)。

と同様の方法で各レンズに対して一つの頂点に対応する簡易なメッシュモデルを推定し、そのモデルに取得された多視点画像を貼り付けることにより自由視点画像を合成する。

Wilburn ら [21, 93, 94, 95] は、動的な light field の撮影のみならず、さまざまな特殊効果の実現（高解像度撮影，高フレームレート撮影，特殊開口合成など）を目的とし，図 2.15 に示されるカメラアレイシステムを開発した．特殊開口合成では，障害物をぼかして背後の物体を鮮明に合成したり，動きぼけ（開口時間が長いことに起因）と焦点ぼけ（開口が空間的に大きいことに起因）のトレードオフ関係を解消する例が示された．そのほかにも興味深い特殊効果を実現する方法が具体的に示され，多眼カメラシステムの用途の広さが明らかにされた．

2.5 多次元信号処理から見た自由視点画像合成

2.3.1 で述べたように，イメージベーストレンダリングは信号処理の過程として考えることができる．すなわち，入力画像群の取得は多次元信号空間（plenoptic function [17]）のサンプリングに相当し，自由視点画像の合成は，あらかじめサンプリングされた多次元信号空間のリサンプリングと考えることができる．したがって信号処理的な解析によって，光線情報を過不足なく取得

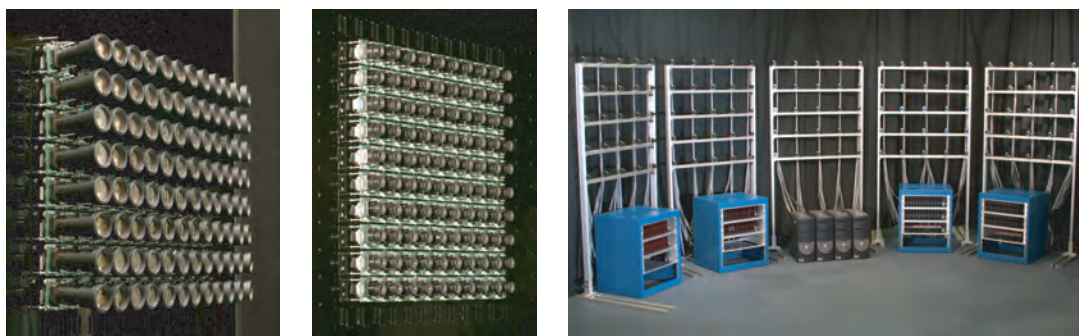


図 2.15: Wilburn らのカメラアレイ：用途に応じて様々な構成を取る（文献 [95] より引用）。

するための条件を求めたり，画像合成の過程を説明することが可能である．

本節では，plenoptic function の四次元表現である light field [23, 24] を対象として議論する．簡単のため， (s, u) 部分空間について考える．図 2.16 に示したように， s はカメラの位置，各カメラの u は画素位置を表す．自由視点画像の合成は，自由な位置にレンダリングカメラを置き，レンダリングカメラの中心に到達する光線を逐一読み出すことによって行う．このとき，実際にサンプリングされていない光線については，適切に補間合成を行う必要がある．この補間は，被写体の同一の点から発している光線を参照して行うのが望ましい．そこで light field rendering では，対象とする空間を一枚の平面（focal plane）で近似し，すべての光線が focal plane 上から発していると仮定して補間処理を行う．この focal plane を置く位置は，画像合成の際に自由に決めることができる．例えば図 2.16 において，破線で示されている光線を得るためには，図中に実線で示されている近隣の光線を用いて補間処理が行われる．

入力画像の取得密度が十分高い（図 2.16 においてカメラ間隔 Δs が十分に小さい）場合には，この手法でシーン全体を鮮明に合成することができる．しかし実用上はサンプリング密度が不十分である場合が多い．その場合，合成画像においては，focal plane 付近の物体は鮮明に合成されるが，focal plane から離れた物体はぼけや ghosting が発生することが知られている [96, 97]．では，シーン全体を十分な品質で合成するためには，入力画像のサンプリング密度をどのように設定したらよいのだろうか．あるいは，合成画像における物体の見え方（ぼけや ghosting）は，画像合成の際の補間の手法によってどのように変わるのだろうか．本節では，これらの疑問に答える信号処理理論について解説する．

2.5.1 Plenoptic Sampling [53]

Chai ら [53] は，光線情報の特性を周波数領域で考察し，最大カメラ間隔および奥行き仮定の最適値を導いている．彼らは，光線情報を扱いやすくするため，1) 非ランバート反射，および 2) オクルージョンの影響が無視できる，という仮定を置いている．

図 2.17 に光線信号の一例を示す．左は (s, u) 空間における信号である． s 軸がカメラの水平位置， u 軸が各カメラの水平ラインに対応する．このパターンはコンピュータビジョンの分野では EPI（epipolar plane image）と呼ばれる．直線状の軌跡は，それぞれ被写体の一点に対応していて，軌

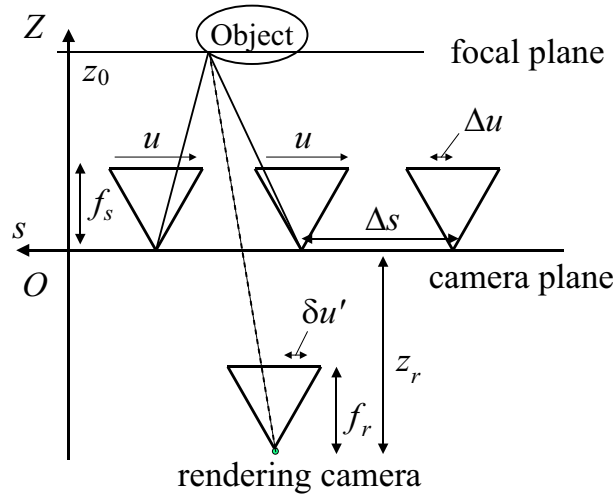


図 2.16: パラメータ設定

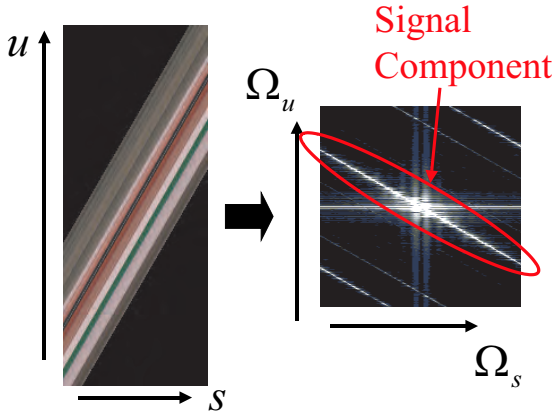


図 2.17: 光線信号の解析

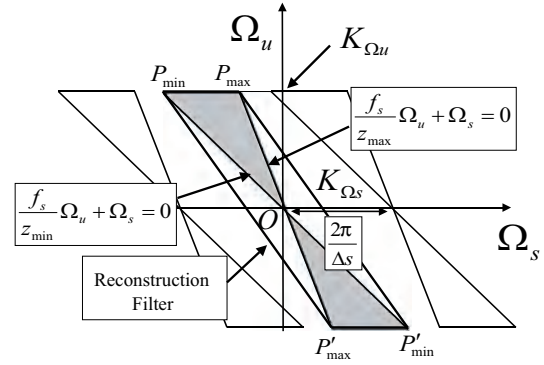


図 2.18: Plenoptic sampling 理論

跡の傾きはその点の奥行きに対応している．この信号をフーリエ変換した結果が右の図 ((Ω_s, Ω_u) 空間，原点は中央) である．EPI 上での軌跡と直交する方向に，直線状に強い信号成分が出ていくことが分かる．その他の成分は，離散化に伴う折り返し雑音および，画像の境界に相当する周波数成分（方形窓関数の畳み込みが，水平・垂直方向にリングングノイズとして現れる）として説明できる．

このように光線信号の周波数特性は非常に偏ったものであり，Chai らの理論はこの点に着目したものである．図 2.18 に Chai らが導出した光線信号の周波数特性を示す．

- まず，対象とする空間の奥行きが $z_{\min} \leq Z \leq z_{\max}$ で定められるとする．このとき周波数領域での信号成分は，下記の 2 本の直線で挟まれた領域に限定される．

$$\frac{f_s}{z_{\min}} \Omega_u + \Omega_s = 0, \quad \frac{f_s}{z_{\max}} \Omega_u + \Omega_s = 0 \quad (2.1)$$

- さらに, u 軸方向の周波数の最大値を $K_{\Omega_u}(= 2\pi K_{f_u})$ とする. この値は, シーンの模様の複雑さやカメラの解像度によって決まるものである.

このとき, 信号成分は図 2.18 の影をつけた領域となる.

離散信号では, $2\pi/\Delta s$ の周期で折り返し成分が発生する. 折り返し成分が元の信号と重なり合わない条件から, 最大カメラ間隔が下記のように求められる.

$$\Delta s \leq \Delta s_{\max} = \frac{1}{K_{f_u} f_s (1/z_{\min} - 1/z_{\max})} \quad (2.2)$$

また連続信号の再構成 (信号値の補間) には, 図中に $P_{\min} P_{\max} P'_{\min} P'_{\max}$ で示される平行四辺形状のローパスフィルタを用いることになる. この平行四辺形の傾きは, 実は画像合成の際に focal plane を仮定する奥行き z_0 に対応する (傾き: $-f_s/z_0$). z_0 の最適値は下記のように定められる.

$$\frac{1}{z_0} = \frac{1}{2} \left(\frac{1}{z_{\min}} + \frac{1}{z_{\max}} \right) \quad (2.3)$$

2.5.2 Generalized Plenoptic Sampling [98]

Zhang ら [98, 19, 99] の generalized plenoptic sampling は, 名前の通り Chai ら [53] の理論を拡張したものである. 両者の差分は主に下記の 2 点である.

1. Chai ら [53] が検討していない, 非ランバート反射とオクルージョンの効果 (詳細は後述) について信号処理的な考察を行った.
2. 正方格子ではないサンプリング格子構造に理論を拡張し, 周波数領域で折り返し信号群を最密充填する方法を示した.

非ランバート反射: 図 2.19 は, 非ランバート反射の影響を示したものである. ある点の反射特性が純粋にランバート反射のみの場合, その点はどんな方向から見ても同じ色に見える. このとき EPI においては, その点に対応する直線状の軌跡の方向には色の変化がない (図 2.17 参照). しかし非ランバート成分が含まれる場合には, 図 2.19 左に示したように, 軌跡の方向に沿って色の変化がある. このとき周波数領域では, 図 2.19 右に示したように, 信号成分に対応する直線状の領域が, 所々“太く”なる効果が現れる. 非ランバート成分が帯域制限されている場合には, この“太さ”を有限に抑えることが可能である. しかし, 鏡や宝石のような“きらきら”した材質の物体は, 高い周波数の非ランバート成分を持つため, この枠組みで扱うのは困難である.

オクルージョン: 図 2.20 は, オクルージョンの影響を示したものである. 被写体の構図は, 奥の平面を手前の平面が隠している, というものである. 図 2.20 左の空間領域では, 二枚の平面それぞれに対応して, 傾きが異なる二つの直線状の軌跡が重なり合うパターンが表れている. 右の周波数領域では, それぞれ平面に対応する二本の信号成分が出ている. 特に丸をつけた領域 (奥の平面に対応) をよく見ると, 手前の面の信号成分と同じ角度の方向 (約 -45° の方向) にもリリングが広がっていることが分かる. これは, 手前の平面に相当するマスキング関数の畳み込みとして説明することができる. ただし, オクルージョンの影響を厳密に知るためには, 被写体の構造を正確に推定する必要がある.

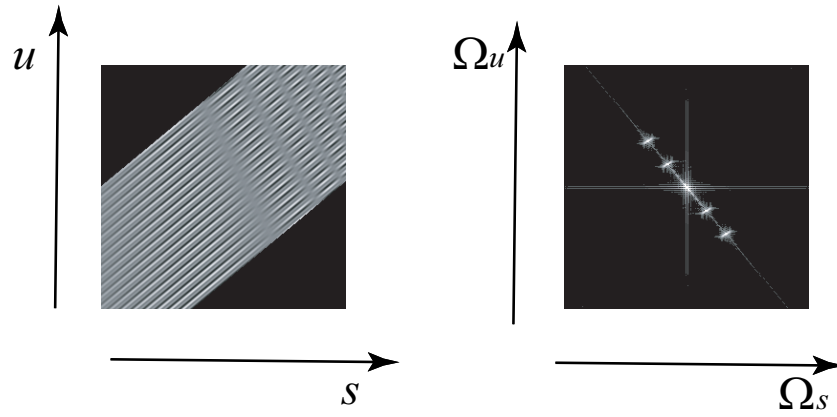


図 2.19: 非ランバート成分を含む光線情報（文献 [98] より引用）．

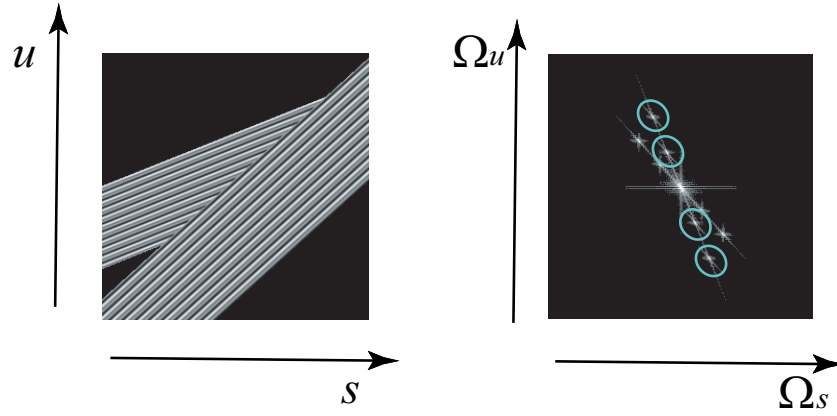


図 2.20: オクルージョンを含む光線情報（文献 [98] より引用）．

折り返し信号の最密充填：ここでは，図 2.21 のように， (s, u) 空間におけるサンプリンググリッドが行列 V で定められる場合を考える．そのとき，周波数領域 (Ω_s, Ω_u) での折り返し信号は，行列 V^T で定められるグリッド上に発生する²．したがって，行列 V とサンプリング密度を適切に選択すれば，図 2.22 のように，折り返し信号を最密充填することが可能である．この場合には，Chai ら [53] の理論による条件と比較して，最大 2 倍の充填率を得られるため，同一品質においてサンプリング密度を半分にすることが可能である．ただし，信号成分の周波数サポートと同じ形の通過域を持つように，再構成フィルタを適切に設計することが条件となる．上記の議論は，理論的には美しいものであるが，現段階では，非正規グリッドでのサンプリングや，再構成フィルタの設計に関して，実用に供するための議論が不十分である．

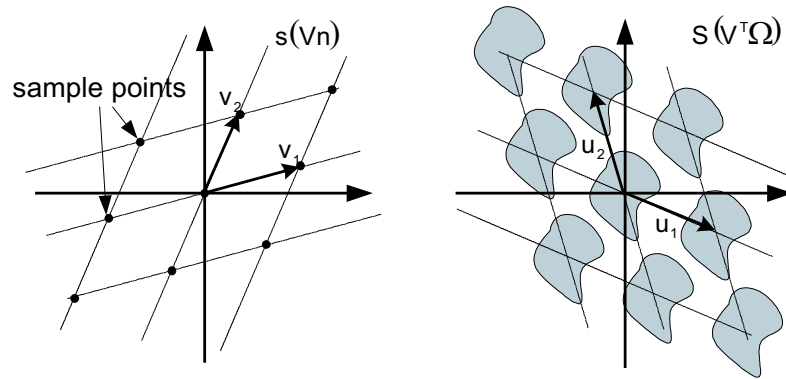


図 2.21: 左：空間領域におけるサンプリング格子の構造．右： 周波数領域における繰り返しの格子の構造（文献 [98] より引用）．

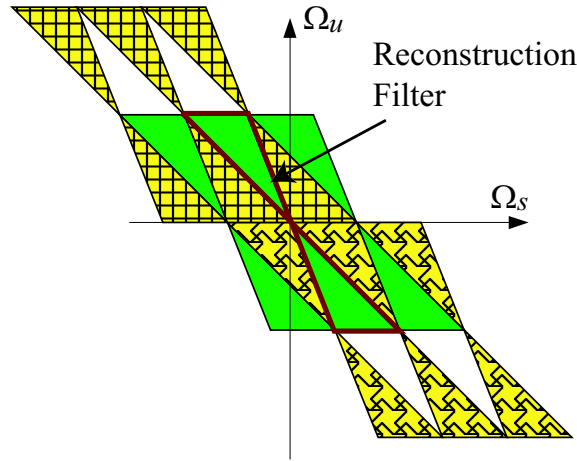


図 2.22: スペクトラムサポートの最密充填（文献 [98] より引用）．

2.5.3 再構成フィルタと合成画像

ここでは，光線のサンプリング密度が不十分で，plenoptic sampling [53] で定式化された条件が満たされていない場合を考える．このとき，周波数領域では，信号成分が折り返し成分と重なり合っている．このような条件下では，信号処理的な意味での正しい信号の再構成を行うことは不可能である．しかし，再構成フィルタ（光線信号の補間を行うフィルタ）の形状を変化させることにより，合成画像にさまざまな効果を与えることができる．ここでは，主に Stewart ら [100] の考察に基づいて，図 2.23 に示した三種類の再構成フィルタを比較する．

- (a) **normal filter**: 最も基本的な再構成フィルタ．サンプリング条件が満たされている場合には，理論上最適なフィルタとなり得る [53]．しかし，図 2.23 においては，エイリアシング雑音を含む再構成となっている．

²plenoptic sampling[53] では， V は対角行列であったため，折り返し信号は，水平，垂直方向に発生していた．

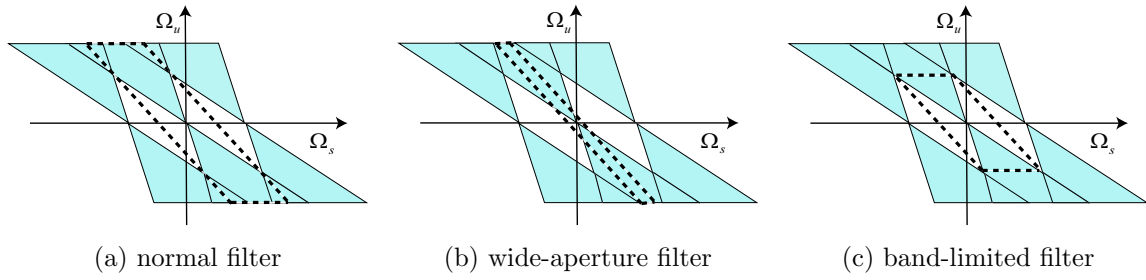


図 2.23: 再構成フィルタの形状



(a) wide-aperture filter

(b) band-limited filter

図 2.24: 再構成フィルタと合成画像（文献 [100] より引用）。

(b) **wide-aperture filter**: normal filter を横方向 (Ω_s 軸方向) に縮小することにより、エイリアシング成分を低減する。図 2.24(a) に示したように、一般のカメラで開口（絞り）を大きくして撮影したような効果が得られ、特定の奥行きのみが鮮明になり他の部分が強くぼける。非ランバート特性などの view dependent な特性の再現性が低いという欠点を持つ。

(c) **band-limited filter**: normal filter を、縦方向 (Ω_u 軸方向) に縮小することによりエイリアシング成分を低減する。図 2.24(b) に示したように合成画像は単純に解像度が低下するためぼける。実装する場合には、normal filter を用いて再構成された画像に対して、ローパスフィルタリングを施すという手順になることが多い。

このように、周波数領域における再構成フィルタの議論は、画像合成の過程や結果をよく説明する。

2.6 本研究の位置づけ

本論文では、カメラアレイによる多眼画像を入力として高品質な自由視点画像を合成する手法について検討する。この検討は、画像群の取得から自由視点画像合成までを実時間化することで、遠隔地通信のようなリアルタイムアプリケーションへの適用を目指すものである。したがって、本

研究は、2.2 で述べた空間共有通信の文脈に位置づけることが可能である。

2.3 では、幾何情報と画像情報のトレードオフ関係について述べた。写実的な画像合成を実現するためには、入力画像の取得密度をできるだけ高めることが望ましい。したがって、本論文では、数十–数百台のカメラで構成されるカメラアレイによる入力を用いることを想定する。しかし、実際のシステムにおいて入力画像の取得密度を高めるのは、カメラの物理的な大きさ、システム構築のコスト、および処理するデータ量の点から限界が生じる。したがって、本論文では取得された画像群を元にシーンの幾何情報を推定し、その情報を画像合成に用いることで画像数の不足を補うことを検討する。図 2.7 に示されるように、このときに必要とされる幾何情報は、画像の数に反比例する。したがって、数十–数百台のカメラ入力を用いる場合には、数台のカメラ入力のみを用いる場合と比較してはるかに簡素な幾何情報で充分であり、必要な計算量をより低く抑えることができる。このように、本論文における検討は、システムの規模や計算量まで含めて、上記のトレードオフ関係を考慮に入れたものである。

2.4 では、多視点画像の取得から自由視点画像の合成までを一連の流れとして捉える観点から、動空間を対象とした自由視点画像合成システムの実例を挙げた。本研究では新たなシステムの開発は行わないが、三次元視覚情報処理のアプリケーションのあり方を思い描き、生じ得る技術上の問題を予測する点で、これらの例は多くの示唆を与える。2.5 では、自由視点画像合成という問題を多次元信号処理の観点から解析する研究例を紹介した。多視点画像の取得は光線情報のサンプリング問題であり、シャノンのサンプリング定理と同様の考察が可能であること、また画像合成は周波数領域における低域フィルタリングとしてモデル化でき、このモデルにより合成画像に表れる現象をよく説明できることを示した。本論文は、これらの理論的考察を発展させ、形状推定という問題へと応用する、新たな展開を導入するものである。

参考文献

- [1] 苗村 健, “光線記述に基づく空間符号化と空間共有メディアに関する研究”, 東京大学大学院工学系研究科, 電子工学専攻, 博士論文 (1996).
- [2] 大越孝敬, “三次元画像工学”, 産業図書 (1972) .
- [3] 大越孝敬, “三次元画像工学”, 朝倉出版 (1991) .
- [4] 磯野春雄, 安田 稔, 竹森大祐, 金山秀行, 山田千彦, 千葉和夫, “多眼式メガネなし 3 次元テレビジョン”, 信学論, Voi. J76-C-II, No. 1, pp. 16–23 (1993)
- [5] 増谷 健, 濱岸五郎, 坂田政弘, 山下敦弘, 中山英治, 井上益孝, “新イメージスプリッタ方式メガネなし 3D ディスプレイ”, 3 次元画像コンファレンス, pp. 90–96 (1996)
- [6] S.A. Benton and H. Yoshikawa, “Recent Experiments in Holographic Video Imaging,” 3D Forum: The Journal of Three Dimensional Images, 5-4, pp. 34–39 (1991).

- [7] S.A. Benton, "The Second Generation of the Mit Holographic Video System," TAO First International Symposium on Three Dimensional Image Communication Technologies, pp. S-3-1 (1993)
- [8] 洗井 淳, 星野春男, 岡野文男, 湯山一郎, "屈折率分布レンズを用いたインテグラルフォトグラフィ撮像方式の基礎検討", 映情学誌, Vol. 51, No. 11, pp.1927–1934 (1997)
- [9] 原島 博, "三次元画像符号化に向けて", 画像符号化シンポジウム (PCSJ92), pp. 9–12 (1992).
- [10] 原島 博, "三次元統合画像通信に向けて", ディスプレイ アンド イメージング, Vol. 1, pp. 121–131 (1992).
- [11] 原島 博, 藤井俊彰, 苗村 健, "三次元統合画像通信の構想", 信学技報, HC92-58, (1993).
- [12] 原島 博, "三次元統合画像通信への道", 3次元画像コンファレンス, pp. 19–24 (1994).
- [13] 藤井俊彰, 原島 博, "構造抽出に基づく三次元画像符号化", 画像符号化シンポジウム (PCSJ94), pp. 23–24 (1994)
- [14] 藤井俊彰, "三次元統合画像通信の基礎検討", 博士論文, 東京大学大学院工学系研究科 電子工学専攻 (1994)
- [15] 藤井俊彰, 金子正秀, 原島 博, "光線群による3次元空間情報の表現とその応用", テレビジョン学会誌 Vol. 50 No. 9 pp. 1312–1318 (1996).
- [16] H.-Y. Shum, S.B. He, and S.-C. Chan, "Survey of Image-Based Representations and Compression Techniques," IEEE Trans. Circuits and Syst. for Video Tech., Vol. 13, No. 11, pp. 1020–1037 (2003).
- [17] E.H. Adelson and J. Bergen, "The Plenoptic Function and the Element of Early Vision," MIT Press, Computer Models of Visual Processing, pp. 3–20 (1991).
- [18] C. Zhang and T. Chen, "A Survey on Image-Based Rendering-Representation, Sampling and Compression," Eurasip Signal Processing: Image Communication, Vol. 19, No. pp. 1–28 (2004).
- [19] C. Zhang and T. Chen, "Spectral Analysis for Sampling Image-Based Rendering Data", IEEE Trans. Circuits and Syst. for Video Tech., Vol. 13, No. 11, pp. 1038–1050 (2003).
- [20] L. McMillan and G. Bishop, "Plenoptic Modeling: an Image-Based Rendering System," Proc. ACM SIGGRAPH 95, pp. 39–46 (1995).
- [21] B. Wilburn, M. Smulski, H.-H.K. Lee, and M. Horowitz, "The Light Field Video Camera," Proc. SPIE Electronic Imaging 2002, Media Processors 2002, Vol. 4674, pp. 29–36 (2002).

- [22] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera,” Eurographics Workshop on Rendering, pp. 1–10 (2002).
- [23] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH 96, pp. 31–42 (1996).
- [24] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and Michael F. Cohen, “The Lumigraph,” Proc. ACM SIGGRAPH 96, pp. 43–54 (1996).
- [25] H.-Y. Shum and L.-W. He, “Rendering with Concentric Mosaics,” Proc. ACM SIGGRAPH 99, pp. 299–306 (1999).
- [26] A. Lippman, “Movie Maps: an Application of the Optical Videodisc to Computer Graphics,” Proc. ACM SIGGRAPH 80, pp. 32–43 (1980).
- [27] G. Miller, E. Hoffert, S.E. Chen, E. Patterson, D. Blackketter, S. Rubin, S.A. Aplin, D. Yim, and J. Hanan, “The Virtual Museum: Interactive 3D Navigation of a Multimedia Database,” The Journal of Visualization and Computer Animation, Vol. 3, No. 3, pp. 183–197 (1992).
- [28] D.G. Ripley, “DVI—a Digital Multimedia Technology,” Communication of the ACM, Vol. 32 No. 7, pp. 811–822 (1989).
- [29] S.E. Chen, “QuickTime VR—an Image-Based Approach to Virtual Environment Navigation,” Proc. ACM SIGGRAPH 95, pp. 29–38 (1995).
- [30] J. Foote and D. Kimber, “FlyCam: Practical Panoramic Video,” Proc. IEEE ICME 2000 (2000).
- [31] U. Neumann, T. Pintaric, and A. Rizzo, “Immersive Panoramic Video,” Proc. ACM Multimedia 2000, pp. 493–494 (2000).
- [32] S. Chen and L. Williams, “View Interpolation for Image Synthesis,” Proc. ACM SIGGRAPH 93, pp. 279–288 (1993).
- [33] S.M. Seitz and C.M. Dyer, “View Morphing,” Proc. ACM SIGGRAPH 96, pp. 21–30 (1996).
- [34] S. Laveau and O.D. Faugeras, “3-D Scene Representation as a Collection of Images,” Proc. IEEE Intl. Conf. on Pattern Recognition, Vol. A, pp. 689–691 (1994).
- [35] S. Laveau and O.D. Faugeras, “3-D Scene Representation as a Collection of Images and Fundamental Matrices,” Technical Report 2205, INRIA, (1994).

- [36] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision,” Cambridge University Press (2000).
- [37] O. Faugeras, “Three-Dimensional Computer Vision: A Geometric View-Point,” MIT Press (1993).
- [38] S. Avidan and A. Shashua, “Novel View Synthesis in Tensor Space,” Proc. IEEE Computer Vision and Pattern Recognition Conf., pp. 1034–1040 (1997).
- [39] L. McMillan, “An Image-Based Approach to Three-Dimensional Computer Graphics,” Ph.D. dissertation, Dept. Comput. Sci., Univ. North Carolina, Chapel Hill, NC (1999).
- [40] M. Oliveira and G. Bishop, “Relief Textures,” Univ. North Carolina, Chapel Hill, NC, Comput. Sci. Technical Report, TR99-015 (1999).
- [41] P. Rademacher and G. Bishop, “Multiple-Center-of-Projection Images,” Proc. ACM SIGGRAPH 98, pp. 199–206 (1998).
- [42] J. Shade, S. Gortler, L.-W. He, and R. Szeliski, “Layered Depth Images,” Proc. ACM SIGGRAPH 98, pp. 231–242 (1998).
- [43] C. Chang, G. Bishop, and A. Lastra, “LDI tree: A Hierarchical Representation for Image-Based Rendering,” Proc. ACM SIGGRAPH 99, pp. 291–298 (1999).
- [44] P.E. Debevec, C.J. Taylor, and J. Malik, “Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-Based Approach,” Proc. ACM SIGGRAPH 96, pp. 11–20 (1996).
- [45] P. Debevec, Y. Yu, and G. Borshukov, “Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping,” Proc. 9th Eurographics Rendering Workshop, pp. 105–116 (1998).
- [46] D.N. Wood, D.I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D.H. Salesin, and W. Stuetzle, “Surface Light Fields for 3D Photography,” Proc. ACM SIGGRAPH 00, pp. 287–296 (2000).
- [47] W.-C. Chen, J.-Y. Bouguet, M.H. Chu, and R. Greszczuk, “Light Field Mapping: Efficient Representation and Hardware Rendering of Surface Light Field,” Proc. ACM SIGGRAPH 02, pp. 447–456 (2002).
- [48] K. Nishino, Y. Sato, and K. Ikeuchi, “Eigen-Texture Method: Appearance Compression Based on 3D Model,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR’99), Vol. 1, pp. 618–624 (1999).

- [49] 西野 亘, 佐藤いまり, 佐藤洋一, 池内克史, “Eigen-Texture 法 : 復号現実感のための 3 次元モデルに基づく見えの圧縮と合成,” 信学論, Vol. J82-D-II, No. 10, pp. 1793–1803 (1999).
- [50] S.B. Kang, R. Szeliski, and P. Anandan, “The Geometry-Image Representation Tradeoff for Rendering,” Proc. Intl. Conf. on Image Processing, Vol. 2, 2000, pp. 13–16 (2000).
- [51] B. Heigl, R. Koch, M. Pollefeys, S. Denzler, and L. Van Gool, “Plenoptic Modeling and Rendering from Image Sequences Taken by Hand-Held Camera,” in Die Deutsche Arbeitsgemeinschaft für Mustererkennung, pp. 94–101 (1999).
- [52] S. Yamazaki, R. Sagawa, H. Kawasaki, K. Ikeuchi, and M. Sakauchi, “Microfacet Billboard-ing,” Proc. Eurographics Workshop on Rendering, pp. 175–186 (2002).
- [53] J.-X. Chai, X. Tong, S.-C. Chany and H.-Y. Shum. “Plenoptic Sampling,” Proc. ACM SIGGRAPH 2000, pp. 307–318 (2000).
- [54] U.R. Dhond and J.K. Aggarwal, “Structure from Stereo – a Review,” IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, No. 6, pp. 1489–1510 (1989).
- [55] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363 (1993).
- [56] A. Laurentini, “The Visual Hull Concept for Silhouette-Based Image Understanding,” IEEE Trans. Pattern Anal. & Mach. Intel., vol. 16, no. 2, pp. 150–162 (1994).
- [57] K. Sato and S. Inokuchi, “Range-Imaging System Utilizing Nematic Liquid Crystal Mask,” Proc. IEEE Intl. Conf. on Computer Vision, pp. 657–661 (1987).
- [58] R. Furukawa and H. Kawasaki, “Uncalibrated Multiple Image Stereo System with Arbitrarily Movable Camera and Projector for Wide Range Scanning,” Proc. IEEE Intl. Conf. on 3-D Digital Imaging and Modeling, pp. 302–309 (2005).
- [59] T. Masuda, Y. Hirota, K. Ikeuchi, and K. Nishino, “Simultaneous Determination of Registration and Deformation Parameters Among 3D Range Images,” Proc. IEEE Intl. Conf. on 3-D Digital Imaging and Modeling, pp. 369–376 (2005).
- [60] J. Yu, L. McMillan, and S. Gortler, “Surface Camera (SCAM) Light Field Rendering,” International Journal of Image and Graphics, Vol. 4, No. 4, pp. 605–625 (2004).
- [61] T. Kanade, P.W. Rander, and P.J. Narayanan, “Virtualized Reality: Constructing Virtual Worlds from Real Scenes,” IEEE Multimedia, Vol. 4, No. 1, pp. 34–47 (1997).
- [62] P.W. Rander, P. J. Narayanan and T. Kanade, “Virtualized Reality: Constructing Time Varying Virtual Worlds from Real World Events,” Proc. IEEE Visualization 1997, pp. 277–552 (1997).

- [63] Narayanan, P., P. Rander, and T. Kanade, “Constructing Virtual Worlds using Dense Stereo,” Proc. IEEE Intl. Conf. on Computer Vision, pp. 3–10 (1998).
- [64] N. Inamoto and H. Saito, “Fly Through View Video Generation of Soccer Scenes,” Intl. Workshop on Entertainment Computing, Workshop Note, pp. 94–101 (2002).
- [65] 稲本菜穂, 斎藤英雄, “多視点スポーツ映像からの自由視点画像合成と提示”, 信学論, Vol. J88-D-II, No. 8, pp. 1693–1701 (2005).
- [66] 松山隆司, 高井勇志, ウ 小軍, 延原章平, “3次元ビデオ映像の撮影・編集・表示”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 4, pp. 521–532 (2002).
- [67] T. Matsuyama and T. Takai, “Generation, Visualization, and Editing of 3D Video,” Intl. Symposium on 3D Data Processing, Visualization and Transmission, pp. 234–245 (2002).
- [68] W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan, “Image-Based Visual Hulls,” Proc. ACM SIGGRAPH 00, pp. 369–374 (2000).
- [69] G. Slabaugh, R. Schafer, and M. Hans, “Image-Based Photo Hulls,” Proc. the First International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 704–708 (2002).
- [70] G.G. Slabaugh, R.W. Schafer, M.C. Hans, “Image-Based Photo Hulls for Fast and Photo-Realistic New View Synthesis,” Real-Time Imaging Vol. 9, No. 5, pp. 347–360 (2003).
- [71] S.M. Seitz and C.R. Dyer, “Photorealistic Scene Reconstruction by Voxel Coloring,” Proc. IEEE CVPR, pp. 1067–1073 (1997).
- [72] K.N. Kutulakos and S.M. Seitz, “A Theory of Shape by Space Carving,” Intl. Journal of Computer Vision, Vol. 38, No. 3, pp. 199–218 (2000).
- [73] M. Li, M. Magnor, H.-P. Seidel, “Hardware-Accelerated Rendering of Photo Hulls,” Proc. Eurographics 2004 (Computer Graphics Forum), Vol. 23, No. 3, pp. 635–642 (2004).
- [74] 北原格, 太田友一, “大規模空間を対象とした自由視点映像生成のための3次元形状復元手法”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 177–184 (2002).
- [75] I. Kitahara, Y. Ohta, “Scalable 3D Representation for 3D Video Display in a Large-scale Space”, IEEE Virtual Reality Conference 2003, pp. 45–52 (2003).
- [76] I. Kitahara and Y. Ohta, “Scalable 3D Representation for 3D Video in a Large-Scale Space,” Presence, Vol. 13, No. 2, pp. 164–177 (2004).
- [77] 苗村 健, 原島 博, “Video-Based Rendering の基礎検討,” 3次元画像コンファレンス 98, pp. 165–170 (1998).

- [78] T. Naemura and H. Harashima, “Real-Time Video-Based Rendering for Augmented Spatial Communication,” SPIE Visual Communication and Image Processing (VCIP ’99), 3653, pp. 620–631. (1999)
- [79] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video-Based Modeling and Rendering of 3D Scenes,” IEEE Computer Graphics and Applications, Vol. 22, No. 2, pp. 66–73 (2002).
- [80] R. Ooi, T. Hamamoto, T. Naemura, and K. Aizawa, “Pixel Independent Random Access Image Sensor for Real Time Image-Based Rendering System,” Proc. IEEE Intl. Conf. Image Processing, Vol. 2, pp. 193–196 (2001).
- [81] 國田 豊, 稲見昌彦, 柳田康幸, 前田太郎, 舘 日章, “相互テレイクジスタンスのための画像構成法”, 日本バーチャルリアリティ学会第3回大会, pp. 203–206 (1998)
- [82] Y. Kunita, Masahiro Inami, Taro Maeda, and Susumu Tachi, “Real-time Rendering System of Moving Objects,” IEEE Trans. Circuits and Systems for Video Technology, Vol. 10, No. 3, pp. 366–375 (2000).
- [83] 國田 豊, 稲見昌彦, 前田太郎, 舘 日章, “多眼カメラを用いた任意視点人物像の実時間生成システム”, 信学論, D-II Vol. J84-D-II No. 1 pp. 129–138 (2001).
- [84] 谷本正幸, “自由視点テレビ”, 三次元映像フォーラム, Vol. 15, No. 3, pp. 17–22 (2001).
- [85] 谷本正幸, “自由視点テレビ FTV”, 映情学誌, Vol. 58, No. 7. pp. 898–901 (2004).
- [86] P.N. Bangchang, T. Fujii, M. Tanimoto, “Experimental System of Free Viewpoint Television,” Proc. of SPIE 5006, Stereoscopic Displays and Virtual Reality Systems X, pp. 554–563 (2003).
- [87] P.N. Bangchang, M.P. Tehrani, T. Fujii, and M. Tanimoto, “RealTime System For Free Viewpoint Television,” 映情学誌, Vol. 59, No. 8, pp. 1191–1198 (2005).
- [88] 中西敦士, 藤井俊彰, 木本伊彦, 谷本正幸, “EPI 上の対応点軌跡を用いた適応フィルタによる光線空間データ補間”, 映情学誌 Vol. 56, No. 8, pp. 1321–1327 (2002).
- [89] C. Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” Proc. Eurographics Symposium on Rendering, pp. 243–254 (2004).
- [90] C. Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” Sketch presentation in ACM SIGGRAPH 04 (2004).
- [91] T. Yamamoto and T. Naemura, “Real-Time Capturing and Interactive Synthesis of 3D Scenes Using Integral Photography,” IS&T/SPIE Electronic Imaging 2004, Stereoscopic Displays and Applications XV, 5291A-16 (2004).

- [92] T. Yamamoto, M. Kojima, and T. Naemura, “LIFLET: Light Field Live with Thousands of Lenslets”, ACM SIGGRAPH 04 Emerging Technologies, etech_0130 (2004).
- [93] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, “High Speed Video Using a Dense Array of Cameras,” Proc. IEEE CVPR, pp. 294–301 (2004).
- [94] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, “Using Plane + Parallax for Calibrating Dense Camera Arrays,” Proc. IEEE CVPR, pp. 2–9 (2004).
- [95] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High Performance Imaging Using Large Camera Arrays,” ACM Trans. on Graphics, Vol. 24, Issue. 3, (Proc. ACM SIGGRAPH 05), pp. 765–776 (2005).
- [96] A. Isaksen, L. McMillan, and S.J. Gortler, “Dynamically Reparameterized Light Fields,” MIT-LCS-TR-778 (1999)
- [97] A. Isaksen, L. McMillan, and S.J. Gortler. “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH 00, pp. 297–306 (2000).
- [98] C. Zhang and T. Chen. “Generalized Plenoptic Sampling,” Technical Report AMP (Advanced Multimedia Processing Lab) 01-06, Carnegie Mellon University (2001).
- [99] C. Zhang and T. Chen, “On Generalized Sampling for Image-Based Rendering Data,” Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP03) (2003).
- [100] J. Stewart, J. Yu, S.J. Gortler, and L. McMillan, “A New Reconstruction Filter for Undersampled Light Fields,” Proc. Eurographics Symposium on Rendering, pp. 150–156 (2003).

第3章

自由視点画像合成における サンプリングと被写界深度の理論

3.1 まえがき

本論文における問題は、イメージベースレンダリングの問題、すなわち「あらかじめ取得された画像群を元に、自由視点画像を合成する」ことである。合成される画像の品質や、合成に要する計算コストは、必然的に、元になる画像データがいかに取得されたか、すなわち入力データのサンプリング条件に依存する。本章の目的は、この依存関係を定式化することにある。ただし、すべてのケースを網羅的に扱うことは容易ではないため、本章では最も基本的なケースを想定した考察を行う。しかし、本章の議論は、より複雑なケースを考える場合にも、基盤となり得るものであると期待される。

本章では、イメージベースレンダリングの最も基本的な形態の一つとして、light field rendering [1] に着目する。この手法では、二次元格子状に配列した多眼カメラによって、入力画像群を取得する。入力画像群は、光線の集合と見なされてインデックス付けされ、光線データベースに格納される。自由視点画像合成は、その視点に到達するすべての光線をデータベースから読出すことによって行われる。ハードウェアのアクセラレーションを利用することにより、リアルタイムで、写実性の高い画像を合成することが可能である。

Light field rendering における画像合成は、あらかじめサンプリングされた光線情報を、リサンプリングするプロセスと考えることができる。このとき、実際にサンプリングされていない光線については、適切に補間合成することが必要である。この補間処理は、対象とする空間全体を1枚の平面 (focal plane) で近似し、所望の光線と同一の点から発する光線を用いることによって行われる。これはすなわち、対象とする空間の奥行きが一定であると仮定することに相当する。そのため、合成画像においては、仮定された奥行きに応じて、“焦点”のような効果が現れる場合がある [2]。すなわち focal plane 付近の物体は鮮鋭に合成される (焦点が合う) が、focal plane から離れた物体にはぼけや2重像が発生する (焦点から外れる)。

では focal plane の前後どれほどの範囲が、“焦点が合っている”状態で描画することが可能なのだろうか。この範囲を定量化できれば、light field rendering による合成画像の品質を直感的に把握することが可能である。そこで本稿では、“焦点が合う”範囲と光線情報のサンプリング条件との関係を定式化する新しい理論 (light field rendering における被写界深度) を提案する。

本理論は、光線情報の信号処理に関して理論的な基盤を与えた plenoptic sampling [3] を拡張することにより導出されるものである。また本稿では、提案理論と、同様に“焦点が合う”奥行きの範囲について定式化した他の二つの理論 (等価被写界深度 [4, 5, 6], 多視点カメラシステムの解像度 [7]) を比較し、提案理論の一般性を明らかにする。さらに画像合成の実験によっても、提案理論の妥当性を確かめる。提案理論によって、イメージベースレンダリングにおけるサンプリング条件の設定や、空間符号化の評価のために、有用な指針が与えられるものと期待される。

3.2 関連研究

本節ではまず、light field rendering のアルゴリズムを説明する。次に、提案理論の信号処理的な基盤となっている plenoptic sampling [3] について説明する。最後に提案理論と同様に、“焦点が

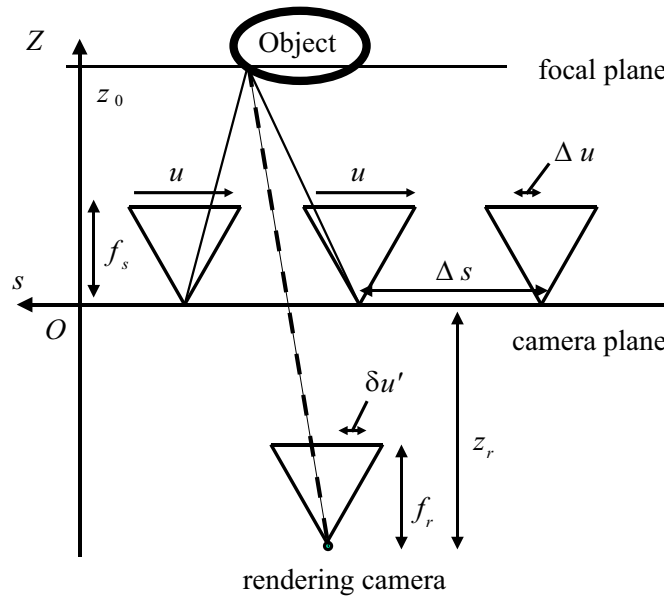


図 3.1: パラメータの取り方

合う”奥行き範囲について考察した二つの理論（等価被写界深度 [4, 5, 6]，多視点カメラシステムの解像度 [7]）を紹介する．なお，これら三つの理論と提案理論との関係については，4 章で詳述する．

3.2.1 Light field rendering のアルゴリズム

Isaksen ら [2] の表記にしたがって，light field rendering のアルゴリズムについて簡単に説明する．光線情報のサンプリングは，2 次元アレイ状に配置されたサンプリングカメラによって行われる．カメラが配置される平面を camera plane と呼び，各カメラの位置は座標 (s, t) で表わされる．各カメラの画像上には (u, v) 座標が定義される．これら四つのパラメータの組 (s, t, u, v) によって，各光線が一意に識別される．

画像合成は，任意の位置にレンダリングカメラを配置し，その位置を通過するすべての光線を，4 次元空間 (s, t, u, v) から逐一読出すことによって行う．必要とする光線が実際に取得されていない場合には，補間処理が必要となる．例えば図 3.1 において，破線で示された光線はサンプリングカメラによって取得されていない．このような場合には，所望の光線（破線）と camera plane との交点を求め，その点の近傍のカメラによって取得された光線群（実線）を用いて，所望の光線が補間合成される．

この補間処理の際には，所望の光線と同一の点から発する光線が使用されることが望ましい．light field rendering では，被写体が存在すると予測される位置に focal plane を配置し，すべての光線が focal plane 上から発していると仮定する．すなわちこれは，対象とする空間の奥行きが一定である，と仮定していることと等価である．

以下本稿では単純のため，4 次元空間 (s, t, u, v) のうち，部分空間 (s, u) を対象として議論を進

めるが, (s, t, u, v) 空間についても同様の議論が成り立つ．以後パラメータの取り方は, 図 3.1 に従うこととする．camera plane に s 軸, 各サンプリングカメラの撮像面に u 軸を定義し, camera plane の位置を原点として, 空間の奥行きを表す軸 Z を取る．focal plane の配置される奥行きを z_0 とする．他のパラメータについては, 適宜説明を加えてゆくことにする．

3.2.2 Plenoptic Sampling

Chai ら [3] は光線情報のサンプリング問題を一般的に扱った, plenoptic sampling を提案している．この理論では, 周波数領域における信号解析に基づき, 最小サンプリング密度と, 理想的な補間のための奥行き仮定の最適値を導いている．

Chai ら [3] は簡単のため, 対象とするシーンについて, 次の2点を仮定している．

1. BRDF (反射特性) モデルは, Lambertian (完全拡散反射) である．
2. オクルージョンの影響が無視できる．

Zhang ら [8] は, これらの仮定が成り立たない場合についても理論を拡張しているが, 本稿では上記の仮定が成り立つ場合について考えることにする．

さらに, 対象とする空間に対する拘束条件として, 次の2点が成り立つとする．

1. 対象とするシーンの奥行きは, $z_{\min} \leq Z \leq z_{\max}$ で定められる．
2. u 軸方向の空間周波数の最大値は, $K_{\Omega_u} (= 2\pi K_{f_u})$ で定められる．

ただし, K_{f_u} は次の式にしたがって定められる．

$$K_{f_u} = \min \left(B_u^s, \frac{1}{2\Delta u}, \frac{1}{2\delta u} \right) \quad (3.1)$$

ここで, B_u^s はシーンテクスチャの最大周波数, Δu はサンプリングカメラの解像度, δu はレンダリングカメラの解像度である．図 3.1 において, レンダリングカメラの解像度は, $\delta u'$ と表記している． δu と $\delta u'$ の関係については, 3.3.3 で後述する．

サンプリングされた信号を, エイリアシングの影響なしに再構成するためには, 周波数領域において, 元信号に対応する成分が隣接する複製成分と重なり合わないようにする必要がある．Chai らは, 光線信号の周波数特性を理論的に解析し, サンプリングカメラの配置間隔 Δs について, 次の条件式を導出した．

$$\Delta s \leq \frac{1}{K_{f_u} f_s (1/z_{\min} - 1/z_{\max})} := \Delta s_{\max} \quad (3.2)$$

この式によって, 最小サンプリング密度 (最大カメラ間隔) が定められる．

さらに Chai らは, 理想的な再構成フィルタの形状を考察し, 奥行きが一定であると仮定する場合, 仮定される奥行き (focal plane の位置) z_0 の最適値は下記のようにして定められることを示している．

$$\frac{1}{z_0} = \frac{1}{2} \left(\frac{1}{z_{\min}} + \frac{1}{z_{\max}} \right) \quad (3.3)$$

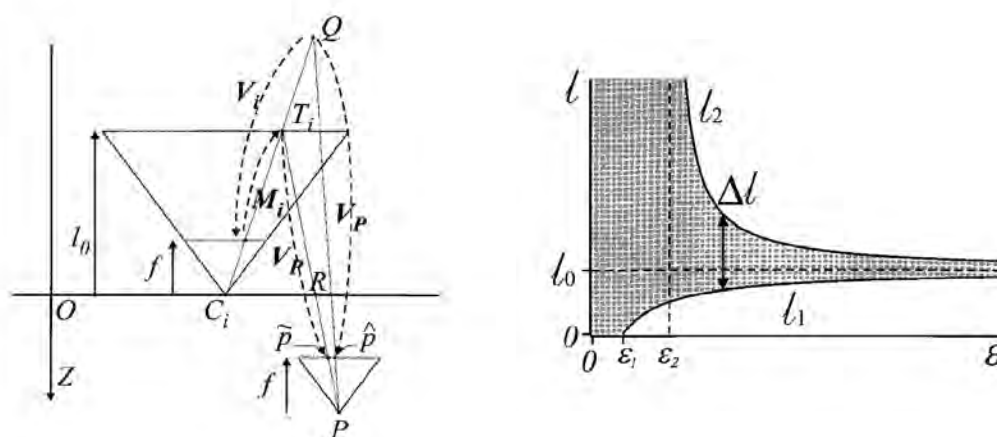


図 3.2: 等価被写界深度 (文献 [6] より引用)

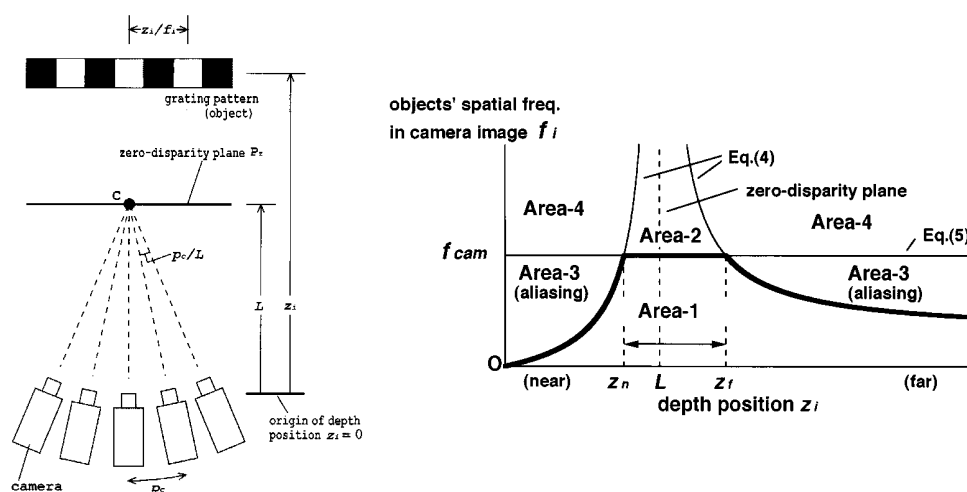


図 3.3: 多視点カメラシステムの解像度 (文献 [7] より引用)

3.2.3 奥行き範囲に対する定式化

國田ら [4, 5, 6] は, 自由視点画像合成において, 対象とする空間の奥行き範囲について定量的な考察を行い, 等価被写界深度を定式化している (図 3.2 参照). 空間中のある点の奥行きが, 仮定された奥行きと一致する場合, その点は合成画像において正しい位置に投影される. しかし, 仮定された奥行きから離れるにしたがって, その点は合成画像において正しい位置から離れた位置に投影される. この理論では, 正しい投影位置と実際に投影される位置との誤差に着目し, この誤差をある閾値内に抑える奥行きの範囲を, 等価被写界深度と定義している. 等価被写界深度とカメラを配置する間隔との関係が示され, カメラ間隔が小さくなる (画像を密に取得する) ほど, 等価被写界深度が広くなることが明らかにされている.

星野ら [7] は, 多視点カメラシステムの解像度 (Resolution of Multi-viewpoint Camera System) を提案している (図 3.3 参照). これは自由視点画像合成ではなく, 3次元ディスプレイのための

多眼撮影を想定した理論である．この理論では，ディスプレイの提示画像において，エイリアシングを発生しないための条件から，空間解像度と対象とする空間の奥行き範囲の関係性を定式化している．結果として，仮定された奥行きから離れるほど，許容される空間周波数の上限値が小さくなる，ということが明らかにされている．

3.3 Light Field Rendering における被写界深度の提案

3.3.1 導出

本節では，light field rendering における“被写界深度”を扱う新しい理論を提案する．本理論は，plenoptic sampling[3] を拡張し，エイリアシングの影響を受けずに合成が可能な奥行き範囲を，定式化したものである．

まず (3.2)(3.3) 式より，それぞれ z_{\max} , z_{\min} を消去することによって，次の2式を得る．

$$\frac{1}{z_{\max}} \geq \frac{1}{z_0} - \frac{1}{2K_{f_u}f_s\Delta s} \quad (3.4)$$

$$\frac{1}{z_{\min}} \leq \frac{1}{z_0} + \frac{1}{2K_{f_u}f_s\Delta s} \quad (3.5)$$

これらの2式をまとめると次のような関係式が得られる．

$$\frac{|Z - z_0|}{Z} \leq \frac{z_0}{2K_{f_u}f_s\Delta s} \quad (3.6)$$

(3.6) 式は，理論的にアンチエイリアスの条件を満たし得る奥行き Z の範囲を示している．すなわち，理想的な再構成フィルタを仮定した場合，この奥行き範囲に相当する光線信号は，エイリアシングの影響を受けずに補間合成することが可能である．

合成画像において，focal plane 付近の物体は鮮明に合成されるが，focal plane から離れた物体にはぼけや雑音が生じる場合がある．このぼけや雑音は，エイリアシングの影響によるものである¹．したがって，エイリアシングの影響を受けない奥行きの範囲を表した (3.6) 式は，直感的には，合成画像において focal plane の前後どれほどの範囲に“焦点が合っている”かを表している，と捉えることができる．そこで，一般のカメラにおける被写界深度とのアナロジーから，(3.6) 式によって定められる領域を light field rendering における被写界深度と定義する．

3.3.2 3通りの表現

(3.6) 式には， Z のほかに四つのパラメータ (Δs , K_{f_u} , f_s , z_0) が含まれる．このうち K_{f_u} と f_s は，両者の積が意味を持つので，独立な変数は三つである．したがって，図 3.4(a)–3.4(c) に示すように，被写界深度に対して3通りの表現が可能である．それぞれ図中の影をつけた領域が，(3.6) 式の条件を満たす領域，すなわち Light field rendering における被写界深度である．それぞれの境界を定める曲線の形は図中に示した通りである．

¹Light field rendering では，ピンホールカメラモデルが仮定されるので，開口やレンズの特性など光学的な要因によるぼけは生じないとする．ここで述べているぼけは原理的に light field rendering に特有なものであることに注意を要する．

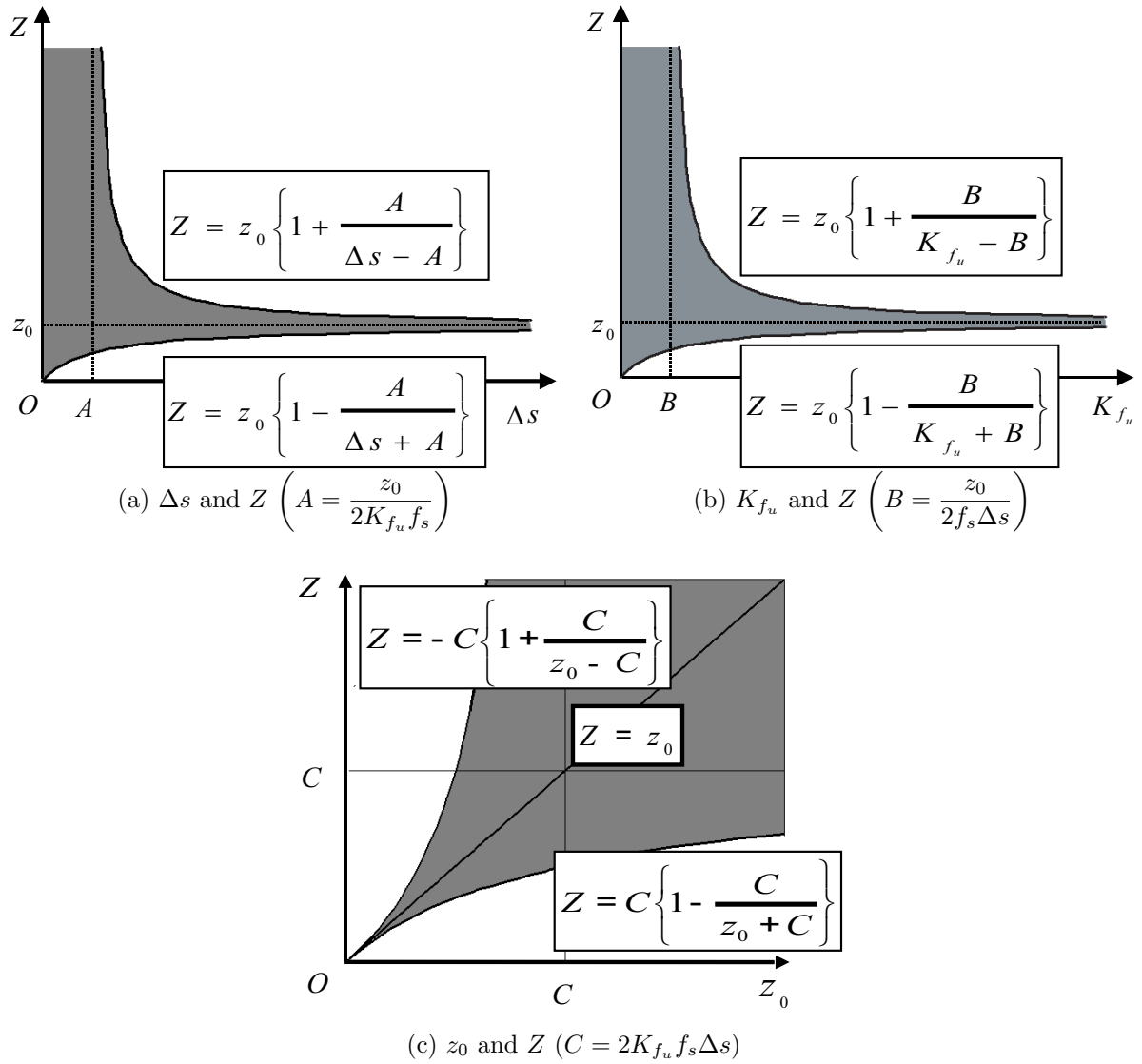


図 3.4: Light field rendering における被写界深度：縦軸は camera plane からの距離，横軸はそれぞれ (a) カメラ間隔，(b) 空間周波数の最大値，(c) focal plane の奥行き

図 3.4(a) は，カメラ間隔 Δs と被写界深度の関係である．カメラ間隔が小さくなる（サンプリングが密になる）ほど，被写界深度が広くなることがわかる．特に $\Delta s \leq A$ では，無限遠が被写界深度に包含される．この図は，必要な被写界深度を得るために，入力画像をどれほどの密度で取得したらよいか，という点で指針を与えるものである．

図 3.4(b) は，空間周波数の最大値 K_{fu} と被写界深度の関係である．グラフの形は図 3.4(a) と同一である．高い空間周波数が必要とされるほど，被写界深度が狭くなることがわかる．この図は，サンプリング，レンダリングの際の画像の解像度の設定に有用である．また，アンチエイリアスのためのフィルタ [1, 2, 7] の設計に利用することができる．図から，低域フィルタによって K_{fu} を抑えれば，エイリアシングの影響を受けない領域（被写界深度）が広がることがわかる．

図 3.4(c) は、仮定される奥行き (focal plane の位置) z_0 と被写界深度の関係を表している。 z_0 が大きくなるほど、被写界深度は広くなる。これは物体が camera plane から遠ざかるほど、隣接するカメラにおける視差が小さくなることに対応している。特に $z_0 \geq C$ では、無限遠が被写界深度に包含される。この図によって、入力画像群を撮影する際に、対象物を camera plane からどれほどの距離に置くかという方針を立てることができる。例えば奥行きに幅がある物体になるほど、camera plane から遠くに置く必要があることがわかる。

3.3.3 レンダリングカメラの解像度についての考察

レンダリングカメラの解像度 (δu) は、(3.1) 式にしたがって空間周波数の最大値 K_{f_u} に影響する。本節では、 δu についての Chai ら [3] の定義を拡張する。この考察は、提案理論を具体的な問題に適用するために不可欠であり、また 4.1 節で提案理論と等価被写界深度 [4, 5, 6] と比較する際にも、その結果を用いることになる。

Chai らの定義した δu は、実際には相対的な値と見なす必要がある。すなわち、実際のレンダリングカメラの 1 画素の大きさを $\delta u'$ とした場合、 δu は $\delta u'$ をサンプリングカメラの撮像面に投影した大きさであると考えられる。これは、空間周波数が u 軸に対して定義されており、 u 軸がサンプリングカメラの撮像面上に張られているからである。

ここで、3.1 の場合について、 δu と $\delta u'$ の関係を求めてみる。レンダリングカメラは自由な方向を向くことができるが、ここでは単純のため camera plane に正対している場合を考える。幾何学的な関係から次の式が導かれる。

$$\delta u = \frac{f_s}{z_0} \frac{z_r + z_0}{f_r} \delta u' \quad (3.7)$$

この式より、 δu には、レンダリングカメラの実際の解像度 $\delta u'$ だけではなく、camera plane からの距離 z_r や仮定される奥行き z_0 が影響することがわかる。これらのパラメータは、自由視点画像合成の際には、ユーザからの要請に応じて動的に変化するものである。したがって、空間周波数の最大値 K_{f_u} は、レンダリングの際に動的に変化するいくつかのパラメータの影響を受けることがわかる。

3.4 他の理論との関係

本章では、第 3.3 章で提案した light field rendering における被写界深度と、他の理論との関係について述べる。

3.4.1 関連理論の導出

第 3.3 章では、plenoptic sampling 理論 [3] から、提案理論 (light field rendering における被写界深度の理論) を導出した。ここでは、図 3.5 に示すように、提案理論から、等価被写界深度 [4, 5, 6] および、多視点カメラシステムの解像度 [7] が導出できることを示す。

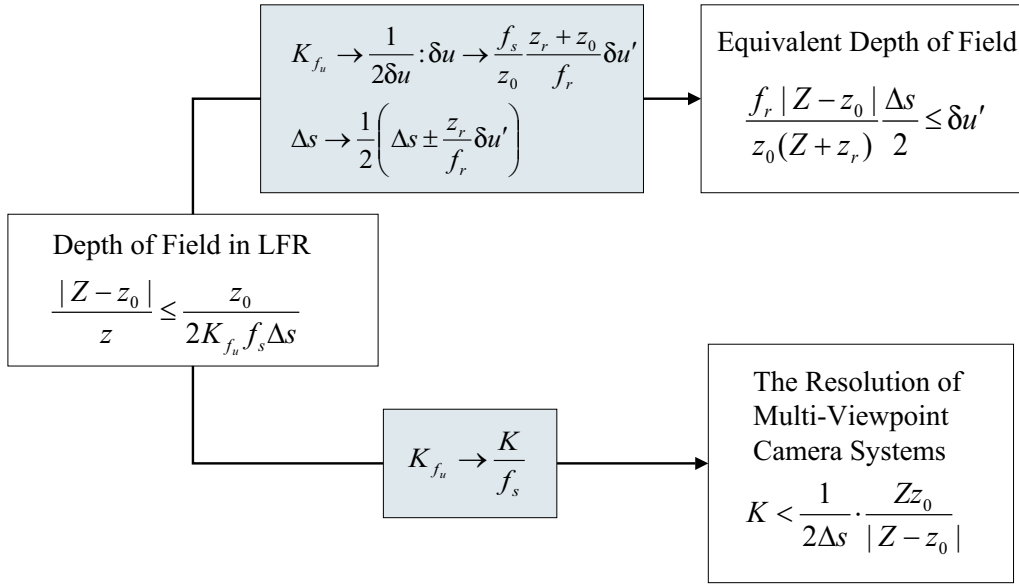


図 3.5: 提案理論からの等価被写界深度および多視点カメラシステムの解像度の導出

等価被写界深度 [4, 5, 6] は，図 3.4(a) と同様に，カメラ間隔と奥行き範囲の関係を示している．図 3.1 の設定に従うと，等価被写界深度は次のように表される．

$$\frac{f_r |Z - z_0| \Delta s}{z_0 (Z + z_r)} \leq \delta u' \quad (3.8)$$

一方，light field rendering における被写界深度において，空間周波数の最大値がレンダリングカメラの解像度によって決定される場合，(3.6) 式に (3.1)(3.7) 式を代入することにより，

$$\frac{|Z - z_0|}{Z} \leq \frac{1}{\Delta s} \frac{z_0 + z_r}{f_r} \delta u'. \quad (3.9)$$

を得る．(3.9) 式に対して，下記の変数変換を施すことにより，(3.8) 式を得ることができる．

$$\begin{aligned} \text{if } Z \geq z_0 : \quad \Delta s &\rightarrow \frac{1}{2} \left(\Delta s + \frac{z_r}{f_r} \delta u' \right) \\ \text{if } Z < z_0 : \quad \Delta s &\rightarrow \frac{1}{2} \left(\Delta s - \frac{z_r}{f_r} \delta u' \right) \end{aligned} \quad (3.10)$$

この変数変換は，“被写界深度”に対する定式化の違いに起因するものである．提案理論では，アンチエイリアスの条件から，信号処理的に元の光線信号が完全に復元できる奥行き範囲を定式化している．一方で等価被写界深度では，合成画像における点の投影位置の誤差（正しい投影位置と，実際の投影位置のずれ）に着目し，この誤差が閾値以内となる奥行き範囲を定式化している．全く異なるアプローチから導かれた両理論が，類似した結果を与えているという事実は興味深い．(3.10) 式の変数変換の詳細については本章の付録で後述する．

表 3.1: 各理論の比較：被写界深度との関係が図示されているパラメータ

	Δs	ΔK_{f_u}	z_0
Plenoptic Sampling[3]	—	—	—
等価被写界深度 [4, 5, 6]	✓	—	—
多視点カメラシステムの解像度 [7]	—	✓	—
Light field rendering における被写界深度	✓ (図 3.4(a))	✓ (図 3.4(b))	✓ (図 3.4(c))

表 3.2: 各理論の比較：空間周波数の要素

	B_u^s	Δu	δu
Plenoptic Sampling[3]	✓	✓	✓
等価被写界深度 [4, 5, 6]	—	—	✓
多視点カメラシステムの解像度 [7]	✓	✓	—
Light field rendering における被写界深度	✓	✓	✓

多視点カメラシステムの解像度 [7] は，図 3.4(b) と同様に，空間解像度と奥行き範囲の関係を示している．図 3.1 の設定に従うと，多視点カメラシステムの解像度は次のように表される．

$$K < \frac{1}{2\Delta s} \cdot \frac{Zz_0}{|Z - z_0|} \quad (3.11)$$

ここで， K は空間周波数である．多視点カメラシステムの解像度は，提案理論と座標系の取り方が異なっている．提案理論では，空間周波数をカメラの撮像面における長さの次元に対して定義しているのに対して，多視点カメラシステムの解像度では，カメラの中心に対する角度の次元に対して定義している．したがって，提案理論の (3.6) 式に対して下記の変数変換をすることにより，(3.11) 式を得ることができる．

$$K_{f_u} \rightarrow \frac{K}{f_s} \quad (3.12)$$

3.4.2 各理論の比較

表 3.1 は被写界深度との関係が図示されているパラメータの比較である．提案理論 (light field rendering における被写界深度) は，plenoptic sampling [3] を拡張したものである．Plenoptic sampling は被写界深度について言及していないが，提案理論では被写界深度を三つのパラメータ (カメラ間隔 Δs ，空間周波数 K_{f_u} ，仮定される奥行き z_0) との関係として図示している (図 3.4(a)–3.4(c))．一方で等価被写界深度 [4, 5, 6] では Δs との関係，多視点カメラシステムの解像度 [7] では K_{f_u} との関係のみを示している．Light field rendering を解析する点では，それぞれの表現形式に意味があり，提案理論はこれら三つの表現形式すべてを示したという点で意義がある．

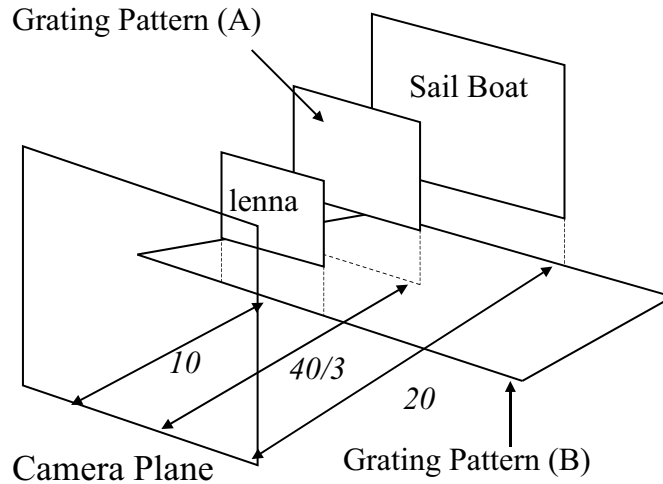


図 3.6: 実験条件

表 3.2 は空間周波数を決める要素の比較である．等価被写界深度 [4, 5, 6] は，空間周波数という概念を持たず，合成画像の解像度のみに注目している．これは， δu を特殊な場合（レンダリングカメラが camera plane に正対している場合）について解いたものに他ならない．多視点カメラシステムの解像度 [7] では，空間周波数の要素として Δu および B_u^s のみを考慮している．この理論は，そもそもイメージベースレンダリングを意図したものではないので，レンダリングカメラの解像度は考慮されていない．一方で提案理論では，plenoptic sampling [3] と同様に， B_u^s ， Δu ， δu の三つの要素すべてが考慮されており，一般性の高い表現となっていることがわかる．

以上の二つの比較から，light field rendering における被写界深度が，plenoptic sampling [3] を拡張すると共に，関連する二つの理論（等価被写界深度 [4, 5, 6]，多視点カメラシステムの解像度 [7]）をまとめ，より一般的に表現したものであることがわかる．

3.5 実験

Light field rendering における被写界深度の理論を検証するため，実験を行った．実験の諸条件の詳細については図 3.6 および表 3.3 に示す．まず，CG によってシーンを構築し，視点位置を等間隔で水平に移動しながら，65 枚の入力画像をキャプチャした．このときの視点位置の間隔は，(3.2) 式により計算された最大カメラ間隔 ($\Delta s_{\max} = 0.075246$) とした．次に，これらの画像群を入力とし，light field rendering のアルゴリズムにしたがって自由視点画像を合成/表示するプログラムを実装した．画像合成のプロセスにおける光線の補間手法としては，実装が簡単な線形補間を採用した²．

²Light field rendering における被写界深度は plenoptic sampling と同様に，理想低域通過型フィルタによる補間を想定している．線形補間は近似的にこの補間を実現する有力な手法であるが，両者の差異については今後議論が必要である．

表 3.3: 実験条件

シーンの条件	$z_{\min}=10, z_{\max}=20, z_0=40/3,$ $B_u^s=\infty$
サンプリングカメラの条件	画角 27° , 画素数 256, $f_s/(2\Delta u)=266.579$
レンダリングカメラの条件	画角 27° , 画素数 512, $z_r=10.0$, $f_s/(2\delta u)=304.662$
空間周波数 K_{f_u}	266.579
最大カメラ間隔 Δs_{\max}	0.0750246

図 3.7(a), 3.7(b) に合成画像の例を示す．本実験において focal plane は, (3.3) 式によって求められる最適な奥行き ($z_0=40/3$) の位置に置かれている．この位置は, 手前の lenna のテクスチャ ($z_{\min}=10$ の位置) と奥の sailboat のテクスチャ ($z_{\max}=20$ の位置) の間に配置されている格子模様の位置と一致する．図 3.7(a) は, カメラ間隔 $\Delta s = \Delta s_{\max}$ の場合の合成画像である．シーン全体が鮮鋭に, “焦点が合った” 状態で合成されていることがわかる．しかし, $\Delta s > \Delta s_{\max}$ となる場合 (入力画像を間引いた場合) には, シーン全体が被写界深度の範囲に収まらなくなる．図 3.7(b) では, focal plane よりも手前や奥の物体において, ぼけや雑音が目につくようになっている．

図 3.7(c) は, (a) と (b) との差分画像である．差分成分を見やすくするためガンマ補正 ($\gamma = 2.0$) を行っている．focal plane の位置には誤差が見られないが, focal plane から前後に離れるほど, 誤差が大きくなる傾向があることがわかる．またこの誤差成分は, カメラ間隔が大きくなるほど, 全体として大きくなる傾向があることが観察された．

図 3.8 は, この誤差成分を数値として表したものである．上に合成画像の PSNR, 下に提案理論に従って計算された被写界深度を示す．対象となるシーンは, $0.75z_0$ から $1.5z_0$ の奥行きの範囲を持っているが, 理論上はカメラ間隔 $\Delta s = \Delta s_{\max}$ において, ちょうどシーン全体が被写界深度に包含される．したがって, PSNR を計算する際, $\Delta s = \Delta s_{\max}$ において合成された画像を基準と考えることにした．

図 3.8 の PSNR は, 対象とするシーンの複雑さや, 画像合成の実装法に依存するものであり, PSNR の値そのものが定量的に意味を持つわけではない．しかし本実験により, カメラ間隔が大きくなるのに伴って, 合成品質が劣化する傾向があるということを確認することができた．提案理論は, カメラ間隔が大きくなるのに伴って, 被写界深度が縮小し, エイリアシング誤差の影響が大きくなることを示しており, 実験結果は理論と一致するものである．

3.6 むすび

本稿では, 自由視点画像合成の基本的な手法である light field rendering に着目し, 元になる光線情報のサンプリング密度と合成品質の関係を定量的に扱うために, 理論的な考察を行った．本

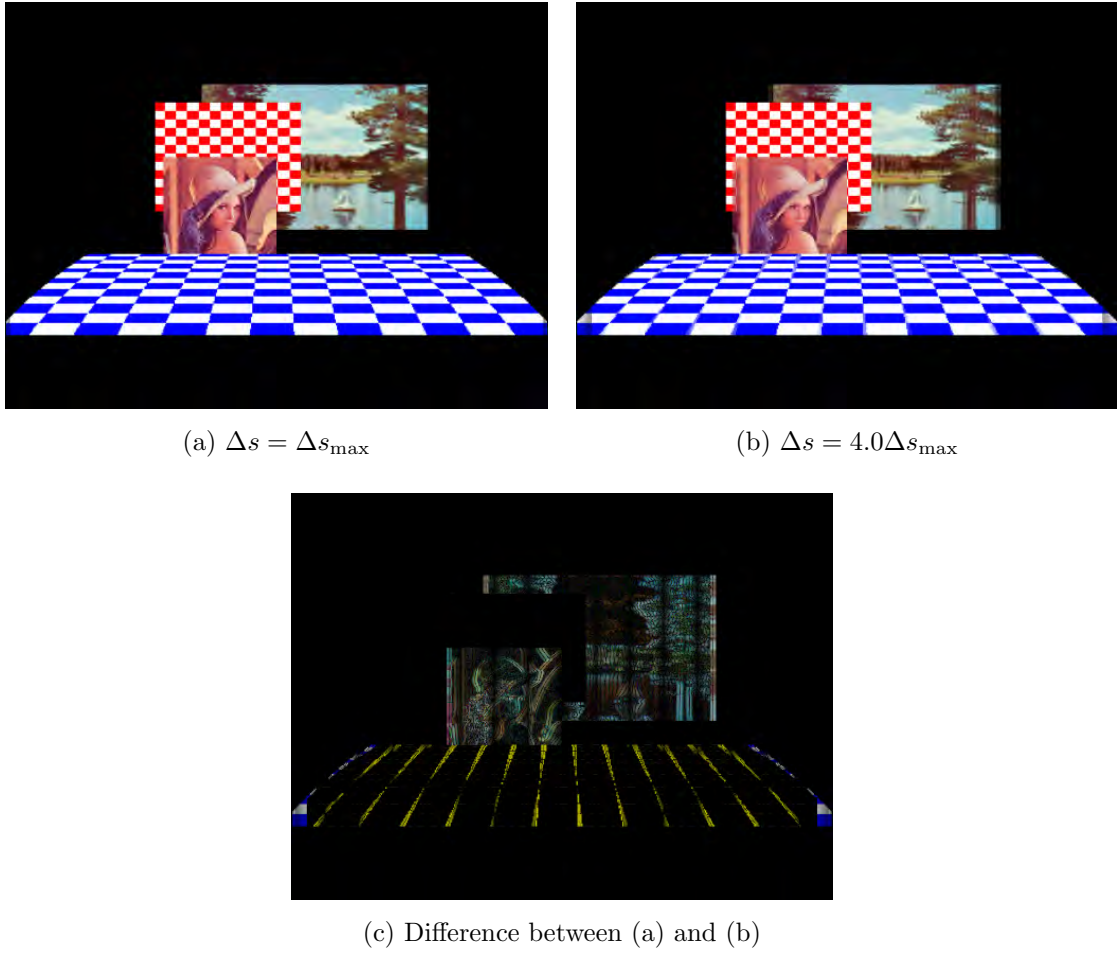


図 3.7: 画像合成：(a) カメラ間隔 $\Delta s = \Delta s_{\max}$ の場合，(b) $\Delta s > \Delta s_{\max}$ の場合 (c) 画像 (a) と画像 (b) の差分

稿で提案する light field rendering における被写界深度は，一般のカメラにおける被写界深度とのアナロジーから，合成画像の品質を直感的に表現したものである．

本理論では，plenoptic sampling [3] に基づき，エイリアシングの影響を受けずに合成することが可能な奥行き範囲を“被写界深度”と定義し，サンプリング密度や他のパラメータとの関係として定式化した (3.6 式)．また空間周波数の要素の一つであるレンダリングカメラの解像度を具体的に表す (3.7 式) ことで，plenoptic sampling [3] をさらに発展させるとともに，等価被写界深度 [4, 5, 6] との関係について考察した．さらに本理論が，全く別のアプローチから導出された等価被写界深度 [4, 5, 6] や多視点カメラシステムの解像度 [7] を一般化したものであることも示した．画像合成の実験によっても，本理論の妥当性を検証した．

本理論の応用例としてまず，自由視点画像合成のための多眼カメラアレイ [9, 10, 11, 12] における，撮影条件の設定を挙げることができる．例えば，水平画素数 256，水平視野角 27° のカメラを 2 cm の間隔で配置してカメラアレイを構成する場合を考える．本理論を適用することにより，

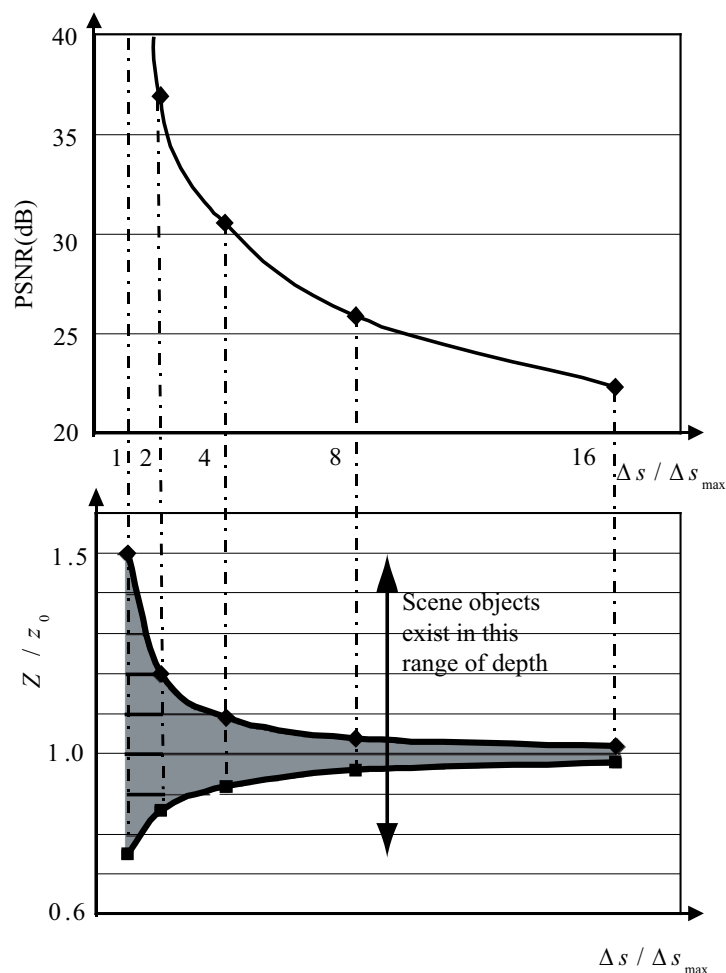


図 3.8: Light field rendering における被写界深度とカメラ間隔についての実験：上の図は合成画像の PSNR，下の図は提案理論に基づいて計算された被写界深度を示している．

被写界深度を最低 1.0 m 確保したい場合、カメラアレイから focal plane までの距離は 2.26 m とすればよいことがわかる．このとき被写界深度は最低でも 1.86 – 2.87 m となる³．

また本理論は、空間符号化 [13] の評価尺度としても利用可能であると期待される．これまでも、多視点画像を相互の相関を利用して符号化する手法について、多くの研究がなされてきたが、そのほとんどは既取得された画像群について、圧縮率や再現性を議論するにとどまっている．しかし、空間そのものを符号化するという観点で考えるなら、画像のサンプリング密度や合成可能な画像の品質まで考慮に入れて、符号化の評価をすることが必要不可欠である．本理論は、そのための基礎を据えたものと位置付けることも可能である．

³最終的な被写界深度はレンダリングのパラメータに依存し、より広い範囲となる場合がある．

参考文献

- [1] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH, pp. 31–42 (1996)
- [2] A. Isaksen, L. McMillan and S.J. Gortler, “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH, pp. 297–306 (2000)
- [3] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, “Plenoptic Sampling,” Proc. ACM SIGGRAPH, pp. 307–318 (2000)
- [4] 國田 豊, 稲見昌彦, 柳田康幸, 前田太郎, 舘 日章, “相互テレイクジスタンスのための画像構成法”, 日本バーチャルリアリティ学会第3回大会, pp. 203–206 (1998)
- [5] Yutaka Kunita, Masahiro Inami, Taro Maeda, and Susumu Tachi, “Real-time Rendering System of Moving Objects”, Proc. IEEE MVIEW’99, pp. 81–88 (1999)
- [6] 國田 豊, 稲見昌彦, 前田太郎, 舘 日章, “多眼カメラを用いた任意視点人物像の実時間生成システム”, 信学論, Vol. J84-D-II, No. 1, pp. 129–138 (2001)
- [7] H. Hoshino, F. Okano, and I. Yuyama, “A Study on Resolution and Aliasing for Multi-Viewpoint Image Acquisition,” IEEE Trans. on Circuits & Syst., Vol. 10, No. 3, pp. 366–375 (2000).
- [8] C. Zhang and T. Chen, “Generalized Plenoptic Sampling”, Technical Report AMP 01-06, Carnegie Mellon University (2001).
- [9] 苗村 健, 原島 博, “Video-Based Rendering の基礎検討”, 3次元画像コンファレンス’98, 5-6, pp. 165–170 (1998).
- [10] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video-Based Modeling and Rendering of 3D Scenes”, IEEE Computer Graphics and Applications, Vol. 22, No. 2, pp. 66–73 (2002)
- [11] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera,” 13th Eurographics Workshop on Rendering, pp. 77–85 (2002).
- [12] B. Wilburn, M. Smulski, H.-H. Keli Lee, M. Horowitz, “Light Field Video Camera,” Media Processors 2002, SPIE, Vol. 4674, pp. 29–36 (2002)
- [13] 高野孝英, 苗村 健, 原島 博, “仮想オブジェクト面を用いた空間符号化,” 信学論, Vol. J82-D-II, No. 10, pp. 1804–1815 (1999)

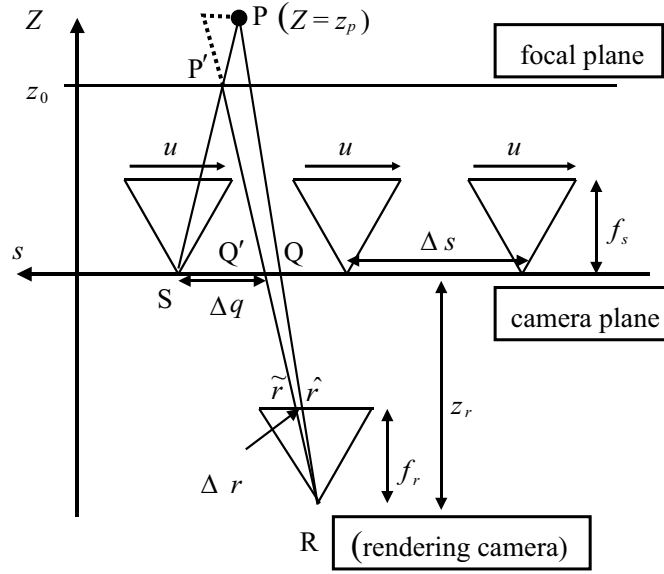


図 3.9: 等価被写界深度の定式化

3.7 付録

本付録では、本文の第3章で提案した light field rendering における被写界深度の理論と、等価被写界深度 [4, 5, 6] を関係付ける (3.10) 式の変換について、幾何学的な説明を行う。(3.10) 式は、

$$\Delta s \rightarrow \frac{1}{2}(\Delta s \pm \frac{z_r}{f_r} \delta u') \quad (3.13)$$

のように、 Δs 軸に対する、定数倍の項 ($1/2$) と、平行移動の項 ($\pm(z_r/f_r)\delta u'$) から成り立つ。それぞれの項について別々に説明を行うこととする。なお、以下では、 $Z > z_0$ の場合について説明するが、 $Z < z_0$ の場合についても同様の議論が成り立つ。

3.7.1 light field rendering における被写界深度

Light field rendering における被写界深度は、特定の条件のもとで、(3.9) 式の形となる。説明の都合上、この式を次のように書き表す。

$$T(Z)\Delta s \leq \delta u' \quad (3.14)$$

ただし、 $T(Z)$ は下記のように定められる。

$$T(Z) = \frac{f_r|Z - z_0|}{Z(z_0 + z_r)} \quad (3.15)$$

3.7.2 等価被写界深度

図 3.9 を用いて、等価被写界深度の定式化について説明する。ここでは R の位置にレンダリングカメラを配置し、合成画像において空間中の点 P ($Z = z_p$) が投影される位置に着目する。等

価被写界深度は，この点の投影位置のずれ（本来投影されるべき位置と，実際に投影される位置のずれ） Δr を求め，そのずれを1画素以内 ($\delta u'$) に抑える条件から導かれる．すなわち下記の条件式が成り立つ場合を考える．

$$\Delta r \leq \delta u' \quad (3.16)$$

対象としているサンプリングカメラの位置を S で表す．直線 $P'R$ と s 軸との交点を Q' とし， $|SQ'|$ の距離を Δq とする．幾何学的な関係から次の式が成り立つ⁴．

$$\Delta r = \frac{f_r |z_p - z_0|}{z_0(z_p + z_r)} \Delta q \quad (3.17)$$

ここで $T'(Z)$ を

$$T'(Z) = \frac{f_r |Z - z_0|}{z_0(Z + z_r)} \quad (3.18)$$

と定めると，(3.17) は次のように表すことができる．

$$\Delta r = T'(z_p) \Delta q \quad (3.19)$$

ここで Δq の最大値を Δq_{\max} とする．(3.16)(3.19) 式より次のような条件式が導かれる．

$$\Delta r = T'(z_p) \Delta q \leq T'(z_p) \Delta q_{\max} \leq \delta u' \quad (3.20)$$

この条件を満たす $Z = z_p$ が，等価被写界深度として定義される．すなわち

$$T'(Z) \Delta q_{\max} \leq \delta u'. \quad (3.21)$$

3.7.3 定数倍の項についての説明

Δq_{\max} は，光線の補間処理の方法に依存する．一般にこの処理は，所望の光線と camera plane との交点を求め，その点の近傍にあるいくつかのサンプリングカメラを参照し，それらカメラによって取得されている光線を用いて行われる（図 3.1 参照）．等価被写界深度ではこの際に，最近傍の一つのカメラのみを参照することを仮定している．したがって (3.21) 式において， $\Delta q_{\max} = \Delta s/2$ が成り立つ（図 3.9 参照）．

しかしここで，近傍の二つのカメラを参照する（線形補間の場合がこれに相当する）場合を考える．この場合には $\Delta q_{\max} = \Delta s$ が成り立つ．このとき (3.14) 式と (3.21) 式の間の変換は，次式のようになる．

$$\Delta s \rightarrow \Delta s \pm \frac{z_r}{f_r} \delta u' \quad (3.22)$$

したがって定数倍の項は，光線の補間処理の際に，カメラを参照する範囲によって決まり，近傍の二つのカメラを参照する場合には1となって無視できることがわかる．

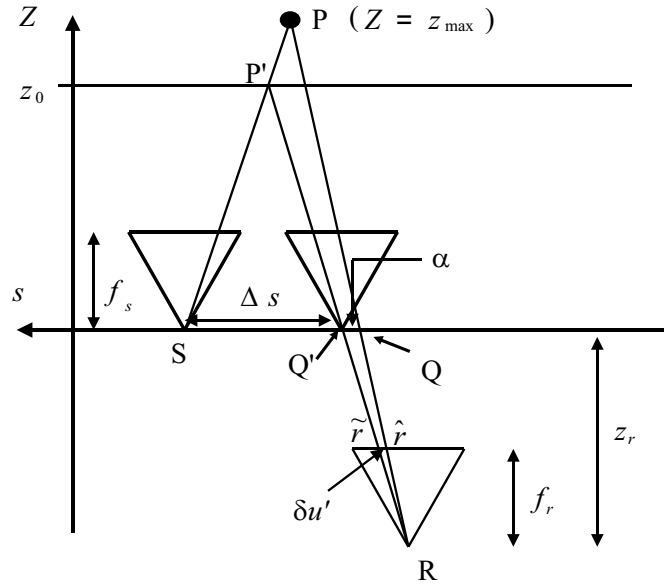


図 3.10: 平行移動の項についての説明

3.7.4 平行移動の項についての説明

ここでは、近傍の二つのカメラを参照する場合について考え、平行移動の項について幾何学的な説明をする。等価被写界深度において、境界条件が与えられる場合を図3.10に示す。このとき点 Q' は対象とするカメラからちょうど Δs 離れた位置（隣のカメラの位置）にあり、 $\Delta q = \Delta q_{\max} = \Delta s$ 、 $\Delta r = \delta u'$ が成立している。

まず、等価被写界深度の定義（(3.21)式）において等号が成り立つ場合を考える。ここでは $Z > z_0$ の場合を考えているので、点 P の奥行きは $Z = z_{\max}$ と表すことにする。

$$\delta u' = T'(z_{\max})\Delta s \quad (3.23)$$

ここで、 PR と s 軸の交点を Q とし $|QQ'| = \alpha$ とする。幾何学的な計算から α は下記のように定められる。

$$\alpha = \frac{z_r}{f_r}\delta u' \quad (3.24)$$

また $|SQ|$ と Δr との幾何学的な関係に注目すると、下記の式が成り立つ。

$$\delta u' = \frac{f_r}{z_r + z_0} \frac{|z_{\max} - z_0|}{z_{\max}} (\Delta s + \alpha) \quad (3.25)$$

ここで、(3.25)式に(3.15)(3.24)式を代入すると、

$$\delta u' = T(z_{\max})\left(\Delta s + \frac{z_r}{f_r}\delta u'\right) \quad (3.26)$$

⁴この関係は、図中に破線で示したように補助線を引くことにより見出される。

結果として, (3.23)(3.26) 式より, 等価被写界深度の境界条件式は, 次のように書き表されることがわかる.

$$\begin{aligned}\delta u' &= T'(z_{\max})\Delta s \\ &= T(z_{\max})\left(\Delta s + \frac{z_r}{f_r}\delta u'\right)\end{aligned}\tag{3.27}$$

これは, 等価被写界深度 ((3.21) 式) が light field rendering における被写界深度 ((3.14) 式) に対する $(z_r/f_r)\delta u'$ の平行移動によって表されることを示している.

第4章

全焦点自由視点画像合成のための 多眼ステレオ法の最適化

4.1 はじめに

多視点画像を入力として、自由視点画像を合成する手法は、イメージベースレンダリング (IBR) と呼ばれる。近年では、動的なシーンを対象とした、インタラクティブな IBR システムが多数研究されている [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]。これらの技術は、遠隔地通信や、放送、シミュレーションなどの分野において、より高い臨場感を実現する有力な手段として、期待を集めている。用途や対象によって、さまざまな形態の IBR システムが検討されているが、本章では、下記のような場合を想定する。

1. 多数 (数十台以上) のカメラを、平面的な 2 次元アレイ状に配列して、入力画像を取得する。
2. 単一のオブジェクトではなく、複数のオブジェクトが存在する被写体空間全体を対象とする。
3. 動的なシーンを実時間で扱うため、オフライン計算は行わないものとする。

上記のようなシステムの例としては、文献 [3, 6, 7, 10] などを挙げることができる。

入力カメラを 2 次元アレイ状に構成した場合の合成画像の品質の限界については、文献 [12] で理論的な体系が与えられている。入力画像の取得密度が十分に高い (カメラの間隔が十分に小さい) 場合には、被写体空間を 1 枚の平面で近似するだけで、良好な画像を合成することができる。一方で、入力画像の取得密度が不十分な場合には、被写体空間を奥行き方向に分割し、複数枚の平面 (奥行きレイヤモデル) で近似することが必要になる。入力画像の取得密度と、必要なレイヤの枚数はトレードオフの関係にある。また、本論文の第 3 章で提案した被写界深度の理論は、各レイヤの前後において、理論上、正しい合成が可能な範囲を表すものであり、文献 [12] を別の方法で表現したものと捉えられる。すなわち、各レイヤを中心とした被写界深度によって、対象とするシーン全体を“覆い尽くす”ことが必要であり、これによって必要なレイヤの枚数が決まる。

実際のシステムにおいて入力画像の取得密度を高めるのは、システム構築のコストの点や、処理するデータ量の点から限界が生じる。そのため、本章では、単一の平面モデルではなく、レイヤモデルを用いる場合を想定する。オフライン計算を前提とする場合には、被写体空間全体を包含するレイヤモデルや、他の形の中間的な 3 次元モデルを、あらかじめ構築しておくことが可能である [13, 14, 15]。一方で、オフライン計算を行わない場合には、求める幾何情報を簡略化したり、幾何情報の精度を犠牲にすることにより、実時間性を確保している例が多い [3, 6, 7, 10]。

本章では、高品位な自由視点画像合成において必要とされる、精度の高い幾何情報を高速に計算する手法を提案する。レイヤモデルを用いる場合に、ある視点位置における画像合成に必要な幾何情報は、その合成画像上の各画素が、どのレイヤに属しているか、という情報である。これは、視点位置に依存した画素単位の奥行きマップ (奥行きの精度はレイヤの枚数に対応する) と等価である。したがって、明示的な 3 次元モデルではなく、視点依存の奥行きマップのみを推定することにより、幾何情報の計算を効率化する。本章の検討は、広い意味での多眼ステレオ法に対して、実時間自由視点画像合成に特化した最適化を行うものとみなされる。より具体的には、提案手法は、多視点画像の color-consistency を用いて photo hull を構築する手法 [16, 17] に対して、計算を必要な範囲に限定することにより効率化を実現するものとみなせる。また、本手法を、グ

ラフィックスハードウェアのアクセラレーションを用いて1台のPC上で実装し、実時間で高品質な自由視点画像を合成できることを確認した。

4.2 関連研究

4.2.1 自由視点画像合成と形状モデル推定

本章では、多数のカメラを用いて入力画像群 (reference view) を取得し、自由視点画像 (target view) を合成することを目的としている。従来のコンピュータビジョン技術は、被写体の形状モデルを構築する空間に応じて、次の3つに分類される。

1. referece view の座標空間で形状モデルを構築する手法
2. 被写体空間に形状モデルを構築する手法
3. target view の座標空間で形状モデルを構築する手法

(1) には、形状復元の分野で古くから研究されている、stereo matching 法 [18] が該当する。基本的な原理は、複数の reference view 間に対応点を探索することにより、視差から距離を逆算するというものである。この手法を用いることにより、個々の reference view の座標空間で、奥行きマップを得ることができる。しかし、一般に、reference view の画素と、target view の画素は1対1に対応しない。これは、両者において、視野に含まれる被写体空間の範囲や被写体相互のオクルージョンの関係が異なるためである。したがって、stereo matching によって得られる個々の奥行きマップは、target view の合成のために十分な情報を持たない。一方で、複数の reference view における奥行きマップの情報を統合すれば、target view の合成に必要な情報を取り出すことが原理的には可能である。この場合の統合処理は、(2) または (3) の座標系で行われることになる。

(2) には、shape-from-silhouette 法 [19] と、shape-from-color-consistency 法 [16, 17] がある。どちらの手法も、被写体空間に定義された各ボクセルが、被写体の内部に含まれるかどうかを判定することにより形状推定を行う。shape-from-silhouette 法では、各 reference view 上の被写体の輪郭を被写体空間に逆投影し、視体積交差の原理により、上記の判定処理を行う。構築される形状は、visual hull と呼ばれる。一方、shape-from-color-consistency 法では、各 referece view 上の各画素の色を被写体空間に逆投影し、対応画素の色の一致度 (color-consistency) を評価する。構築される形状は、photo hull と呼ばれる。一般に、visual hull よりも、photo hull のほうが、計算コストを要するが、より正確な形状が推定できる。これらのモデル構築には、膨大な計算が必要であるため、一般には、オフライン処理を必要とすることが多い¹。しかし、いったんこれらのモデルが構築できれば、これを target view に対して投影することにより、任意の target view の合成に必要な幾何情報を容易に得ることができる。

一方、(3) は、実時間処理を念頭に置いたアプローチであり、被写体のモデルを、target view の座標空間で直接構築する [2, 7, 8, 9, 11, 10]。これは、暗に visual hull または photo hull を推定

¹モデルの解像度の調節や PC クラスタを用いた計算の分散化により、処理時間を短縮し実時間性を確保することも可能である [5]。

表 4.1: 視点依存形状推定手法の比較

	Matusik[2]	Yang[7]	Slabaugh[8]
without silhouette	no	yes	no
per pixel resolution	yes	yes	no
region-based consistency verification	-	no	yes
const. cost to # of cameras	no	no	no

Kitahara[9]	Zhang[10]	Li[11]	proposed
no	yes	no	yes
yes	no	yes	yes
-	yes	no	yes
yes	yes	no	yes

していることと等価であるが，下記の2つの点で，target view に特化した効率化を行うことが可能である．

- モデル推定の範囲：target view の視野角の範囲外や死角になる部分（被オクルージョン領域やオブジェクトの背面）についての情報は不要である．
- 空間解像度：高々target view の画素間隔に対応する解像度で充分である．

このアプローチによって得られるモデルは，特定の target view に特化した（視点位置に依存した）ものであり，新たな view を合成する際には，逐一モデルを構築しなおす必要がある．これは特に，動空間を対象とした実時間処理に適する．

4.2.2 視点依存形状モデル

本節では，オフライン計算を必要としない自由視点画像合成を対象とするため，特に (3) のアプローチに着目する．従来の手法 [2, 7, 8, 9, 11, 10] の比較を表 4.1 に示す．なお，この表では，一般的な意味での分類ではなく，本章の目的に照らして，望ましいと考えられる性質をあげ，各手法がそれらの性質を満たすかどうか（yes/no）を示した．

まず，これら手法は，モデル推定に silhouette のみを用いる手法 [2, 9]，color-consistency のみを用いる手法 [7, 10]，およびその両者を組み合わせる手法 [8, 11] に分類できる．両者を組み合わせる手法では，最初に silhouette を用いて visual hull を求めておくことにより，color-consistency を評価する空間の範囲を限定し，計算コストを抑えるという戦略を取っている．silhouette を用いる場合には，背景差分などを用いて背景とオブジェクトが分離できることが前提となるため，被写体空間全体ではなく，特定のオブジェクトを対象とする用途に適する．また，そのような場合には，入力カメラは，円周上や半球上に，オブジェクトを見込むように配置されることが多い．しか

し、本章では、カメラを平面的な2次元アレイ状に配置し、複数のオブジェクトを含む被写体空間全体を扱うことを想定している。したがって、本章の目的に照らして望ましいのは、silhouetteを用いない手法であると考えられる（表4.1の最初の項目：without silhouette）。

表4.1の2番目の項目には、構築されるモデルの空間解像度を挙げる。本来、このモデルでは、target viewの各画素に対して固有の奥行きが割り当てられることが望ましい（per pixel resolution）。3番目の項目には、領域ベースでcolor-consistencyの評価を行うかどうか（region-based consistency verification）を挙げる。画像のノイズやカメラキャリブレーションの誤差の影響を吸収するためには、各reference viewから1画素のみを読み出して評価するのでは不十分であり、画素集合（領域）を単位とした評価が望ましい。また、誤対応を避ける観点からも、領域ベースの評価が有効である。実時間処理の範囲内で、per pixel resolutionとregion-based consistency verificationの両方を満たすのは困難である。文献[6, 11]では、画素単位での奥行き推定を実現しているものの、領域ベースの処理を行っていないため、結果として得られる画像はノイズの影響を受けやすい。一方で、文献[8, 10]では、領域ベースの処理により、color-consistencyを安定に評価するが、ダウンサンプリングや多重解像度メッシュを用いて、構築するモデルの解像度を低くするアプローチを取っている²。

表4.1の最後の項目は、reference viewの数に対して、ボクセル一つあたりの計算コストが一定であるかどうか、という点である。比較的少ない台数のカメラを用いるシステム[2, 6, 8, 11]では、計算コストが、reference viewの数におおよそ比例して増大するアルゴリズムが用いられている³。一方、数十台以上のカメラ入力を想定するシステム[9, 10]では、reference viewの数に対して、計算コストが一定となるアルゴリズムが採用されている。

上記の分類項目に対する提案手法の特性は下記の通りである（表4.1の最右列参照）。提案手法では、複数のオブジェクトが存在する被写体空間全体を扱うことを想定するため、silhouetteを用いずcolor-consistencyのみを用いて形状推定を行う。また、構築される視点依存モデルの解像度は、合成画像の画素数に対応したものであり、かつ領域ベースの処理により、color-consistencyの評価の安定性も確保する。さらに、数十台以上の大規模なカメラアレイ入力を扱うことを想定するため、カメラの台数に対して計算量が一定となるアルゴリズムを採用する。次章より明らかなように、提案手法を構成する個々の要素技術は決して斬新なものではない。提案手法の意義は、各要素技術を適切に組み合わせ、表4.1に挙げた4つの条件すべてを満たす処理を、実時間で実現していることにある。

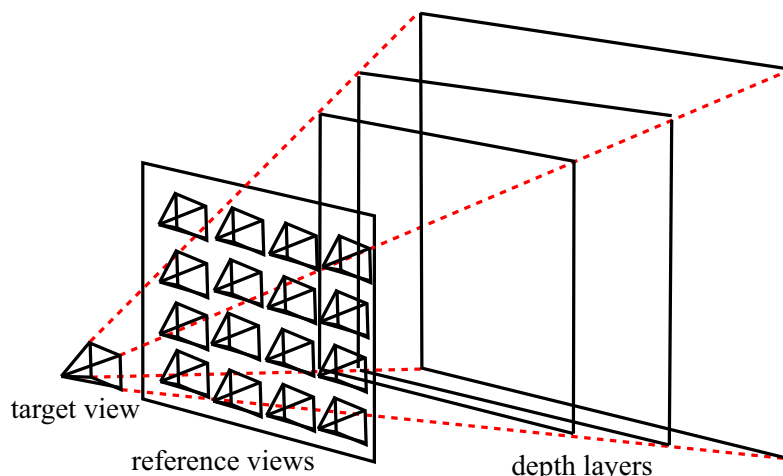


図 4.1: 基本的な配置

4.3 提案手法のアルゴリズム

4.3.1 設定

提案手法では、カメラを2次元アレイ状に配置して、入力画像 (reference view) を取得する場合を想定する。これらのカメラは、事前にキャリブレーションされている (内部パラメータ、外部パラメータが計測されている) ものとする。これにより、カメラ間の個体差やカメラ配置の誤差を吸収できる。提案手法の目的は、オフライン計算を行わず、インタラクティブに自由視点画像 (target view) を生成することである。また、本手法では、単一のオブジェクトではなく、target view の視野に入る被写体空間全体を処理の対象とするため、silhouette 情報を用いないものとする。

画像合成の際には、図 4.1 に示すような、視点位置に依存する奥行きレイヤモデルを利用する。各レイヤは、入力カメラ面と平行に配置するものとする。target view の合成に必要な十分な奥行き情報は、target view の各画素がどのレイヤに属するのか、というものである。これは、視点位置に依存した奥行きマップを得ることと等価である。このとき、target view の死角になる部分や視野角の範囲外についての知識は不要であり、また、target view の解像度を超える精度も不要である。本章では、shape-from-color-consistency 法 [16, 17] のアルゴリズムを基に、target view の合成に特化した、効率的な奥行きマップ推定法を提案する。

本節の構成は下記のとおりである。まず 4.3.2 で、レイヤの設定方法について議論する。4.3.3 では、reference view の数に対して、ボクセル一つあたりの計算コストが一定となるような、color-consistency の評価尺度を導入する。4.3.4 では、color-consistency の評価を安定化するための、効率的な平滑化処理について述べる。4.3.5 では、各画素に対する奥行きの割り当ての方法、4.3.6 では、各画素の色付けの方法について述べる。最後に、4.3.7 で、提案手法の擬似コードを示す。

²Slabaugh[8] の手法で、ダウンサンプリングを用いない場合、処理レートは 0.4 fps 程度であることが報告されている [11]。

³ここで、“おおよそ” 比例と述べたのは、shape-from-color-consistency 法では、visibility のチェックにより、評価の対象から除外される reference view があるためである。

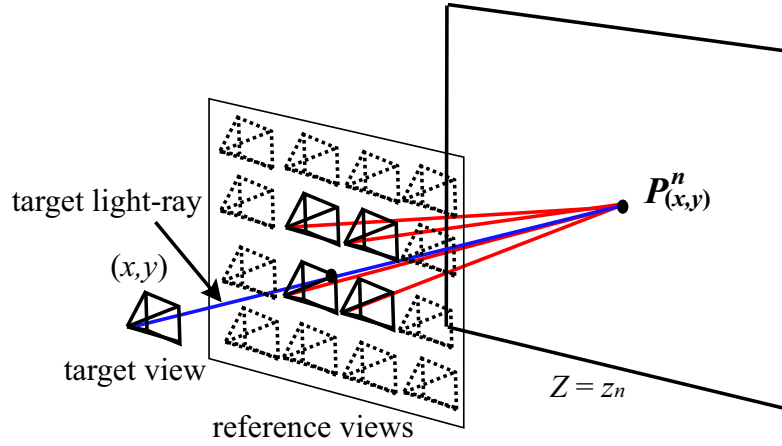


図 4.2: color-consistency の評価

4.3.2 レイヤモデル

入力カメラが、厳密に平行かつ等間隔に配置されている場合には，Chai ら [12] の理論を直接当てはめて，レイヤの配置を決定することができる．提案手法では，厳密なカメラ配置を必ずしも前提とはしないが，入力カメラを 2 次元格子に近い形状に並べた場合には，Chai ら [12] の理論から得られる知見が有用なものであると考えられる．

Chai らによると，まず，合成画像に歪が生じないために必要なレイヤの枚数は，オブジェクトが存在する奥行き範囲，画像（reference view および target view）の解像度，reference view の間隔などによって決定される．特に reference view の間隔と，レイヤの枚数は反比例の関係にある．これは，カメラアレイの面積が一定の場合，reference view の数を増やしたほうが，レイヤの枚数が少なくてよいことを意味する．したがって，少ない台数のカメラ（数台）を用いたシステムでは，数十～百枚程度のレイヤが用いられた [7, 11] が，本章で対象とする数十台以上のカメラを用いた大規模なカメラアレイでは，より少ない枚数のレイヤで充分である．

必要なレイヤの枚数を N ，被写体空間の奥行き（入力カメラ面からの距離）の最大値，最小値をそれぞれ， Z_{\max} ， Z_{\min} としたとき， n 番目のレイヤの奥行き z_n ($n = 1, 2, \dots, N$) は，Chai らの理論に従い，次式で与えることにする．

$$\frac{1}{z_n} = \frac{1}{Z_{\max}} + \frac{n - 1/2}{N} \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) \quad (4.1)$$

この式では，奥行き Z の逆数を等間隔に分割するようにレイヤが配置される．したがって，遠方に行くほどレイヤの間隔が疎になる．また，この式は $Z_{\max} = \infty$ の場合にも適用可能であり，無限遠を包含するような被写体空間を扱うことも原理的に可能である．

4.3.3 局所的な color-consistency の評価

前述したように，target view を合成するためには，target view の各画素が，どの奥行きレイヤに属しているのかという情報を得る必要がある．提案手法では，target view の各画素に対応す

る光線と、各レイヤとの交点において、reference view の color-consistency を評価することにより、この目的を実現する。これは、(合成画像の画素数) × (レイヤの枚数) 個のボクセルを定義し、各ボクセルがオブジェクトに含まれる確からしさを評価していることに相当する。まず、本節では、そのための評価値の計算方法を導入する。

一般に、photo hull を構築する場合には、visibility が保証されるすべての reference view を対象として、color-consistency を評価する [16, 8, 11]。このとき、ボクセル一つあたりの計算コストは、reference view の数にほぼ比例して増大する。しかし、本章では、数十台以上のカメラからなる大規模なカメラアレイを入力として用いることを想定するため、計算コストが、カメラの台数に対して独立であることが望ましい。

そこで、本章では、文献 [10] と同様に、color-consistency の評価を、数個の reference view に限定する方法を採用する。具体的には、図 4.2 に示したように、奥行き推定の対象となる光線 (target view の各画素に対応する光線) に近接する 4 つの reference view のみを評価の対象とする。このとき、ある点で color-consistency を評価するための計算コストは、reference view の数に依存せず、一定である。さらに、計算コストの他にも、次のようなメリットがある。

- オブジェクトの反射特性は、対象とする光線の近傍で、局所的にランバートで近似できればよい。したがって、完全にランバート反射ではないオブジェクトを扱うことが可能である。
- 評価の対象となる点が、近傍の 4 つの reference view からのみ見えていればよい。visibility のチェックを省略しても、オクルージョンの問題が大きく影響することはない。
- 画像合成の際には、見えの視点依存性を保つため、近傍の reference view の光線のみを用いて、対象とする光線の色を補間合成することが多い [20]。提案手法では、consistency の評価を、合成に用いられ得る光線のみに限定する。

以上の内容を、記号を用いて書き下すと、次のようになる。target view の画素の位置を (x, y) で表し、 (x, y) を通過する光線と、 $Z = z_n$ のレイヤとの交点の斉次座標を $p_{(x,y)}^n$ とする (図 4.2 参照)。また、 (x, y) を通過する光線に近接する、4 つの reference view の平面射影を表す斉次行列を、それぞれ、 H_1, H_2, H_3, H_4 とする。 $p_{(x,y)}^n$ に対する評価関数を次のように定義する。

$$E(p_{(x,y)}^n) = \text{consistency} \left(\text{RV}_i \left(H_i p_{(x,y)}^n \right) \right)_{i=1,2,3,4} \quad (4.2)$$

ただし、 $\text{RV}_i(H_i p)$ は、 i 番目の reference view RV_i 上の、座標 $H_i p$ に対応する画素の色を表すものとする。

color-consistency を評価には、対応する画素の SSD (Sum of Squared Difference), SAD (Sum of Absolute Difference), 分散 (または標準偏差) などが用いられるが、本章では最もよく用いられている分散を採用する [16, 8, 11]。すなわち、分散の値が小さいほど、consistency が高いと判断する。各画素は、RGB の 3 成分を持つため、単純にこれらの成分の分散の和を評価値として用いることにする。例えば、R 成分 ($r_i(H_i p_{(x,y)}^n)$) の分散は次のように与えられる。

$$\sigma_r^2(p_{(x,y)}^n) = \frac{1}{4} \sum_{1 \leq i \leq 4} \left(r_i(H_i p_{(x,y)}^n) - \hat{r}(p_{(x,y)}^n) \right)^2 \quad (4.3)$$

ただし, $\hat{r}(p_{(x,y)}^n)$ は, 平均値:

$$\hat{r}(p_{(x,y)}^n) = \frac{1}{4} \sum_{1 \leq i \leq 4} r_i(H_i p_{(x,y)}^n) \quad (4.4)$$

である. σ_g^2, σ_b^2 も同様にして求める. 点 $p_{(x,y)}^n$ に対する評価値は次のように定義される.

$$E(p_{(x,y)}^n) = \sigma_r^2(p_{(x,y)}^n) + \sigma_g^2(p_{(x,y)}^n) + \sigma_b^2(p_{(x,y)}^n) \quad (4.5)$$

4.3.4 評価値画像空間での平滑化

(4.5) 式では, 各 reference view から一画素のみを取り出して評価しているため, 誤対応が生じやすい. また, 画像の雑音やカメラキャリブレーションの誤差の影響を受けやすく, 評価が不安定になりやすい. これらを解決するためには, $H_i p_{(x,y)}^n$ を中心とするウィンドウ領域 (画素集合) に対して評価を行うことが有効であるが [7, 8, 10, 11], 計算量の増大が問題となる. 例えば, 一辺 L 画素のウィンドウ単位でのマッチングを直接行う場合, (4.5) 式の計算を L^2 回行い, さらに得られた L^2 個の評価値の (加重) 平均を取る必要がある.

そこで, 本章では, 文献 [7] と同様の手順により, 各 $p_{(x,y)}^n$ に対する (4.5) 式の計算を 1 回のみとして, ウィンドウマッチングをした場合と等価な効果を得ることとする. まず, $Z = z_n$ のレイヤに対する各画素 (x, y) の評価値 $E(p_{(x,y)}^n)$ を 2 次元に並べる. これを評価値画像と呼び, $E_n(x, y)$ と表記する. この評価値画像を, 次式にしたがってブロック平滑化することにより, reference view 間でウィンドウマッチングをするのと, 等価な効果が得られる.

$$\hat{E}_n(x, y) = \sum_{-W \leq m \leq W} \sum_{-W \leq n \leq W} \frac{E_n(x+m, y+n)}{(2W+1)^2} \quad (4.6)$$

ここで, W は正数であり, $L = 2W + 1$ となる.

さらに, $\hat{E}_n(x, y)$ を, 次のような漸化式を用いて計算することにより, 平滑化の計算のオーダーを $O(L^2)$ から $O(L)$ に削減することが可能である.

$$\hat{E}_n(x, y) - \hat{E}_n(x-1, y) = \sum_{-W \leq n \leq W} \frac{E_n(x+W, y+n) - E_n(x-W-1, y+n)}{(2W+1)^2} \quad (4.7)$$

4.3.5 最小値探索による奥行き決定

文献 [8, 11] の手法では, 閾値処理によって, 各画素の奥行きが決定される. すなわち, 各画素について, レイヤを手前から奥に向かって順に評価してゆき, 評価関数が閾値以下となった場合に, その奥行きにオブジェクトが存在すると判断する (この状態を, 収束と呼ぶことにする). すでに収束している画素については, その後の評価の対象とする必要がないため, 評価のコストを削減することも可能である [8].

閾値を用いたアルゴリズムでは, 適切な閾値の設定が鍵を握る. 閾値が大きい場合, 実際のオブジェクト表面からかけ離れた位置に photo hull が復元されてしまう. 閾値が小さい場合, レイヤをすべて評価しても, 収束しない画素が現れる. 一般的なシーンでは, 領域によって, 反射特性や, 表

面の模様の複雑が異なるため、シーン全体の対して好ましい結果を与えるような color-consistency の閾値を決めることは困難である。

そこで本章では、文献 [7] と同様に、各画素についての奥行き割り当てを、最小値探索によって行うことにする。すなわち、画素 (x, y) の奥行きのインデックス $n(x, y)$ は、次のように求める。

$$n(x, y) = \arg \min_n \hat{E}_n(x, y) \quad (4.8)$$

ここで、 $n(x, y)$ は、target view の位置に応じた、画素単位の奥行きマップとなる。

閾値処理によって構築される photo hull とは異なり、提案手法では、推定されたモデルが真の形状を内包することは保証されない。しかし、すべての画素に対して、必ず一つの奥行きが割り当てられることは保証されるため、結果として「穴」のない合成画像を得ることができる。また、最小値探索では、すべてのレイヤを評価するまで各画素の奥行きを決定できないため、一般の photo hull 構築において行われるような、visibility のチェックが適用できない。しかし、提案手法では、3.3 で述べたように、color-consistency の評価を、対象となる光線に近接する reference view に限定することによって、オクルージョンの影響を低減している。

4.3.6 各画素の色づけ

最後に、推定された奥行き $n(x, y)$ に基づいて、target view $T(x, y)$ の各画素の色を決定する。

$$\begin{aligned} r(T(x, y)) &= \frac{1}{4} \sum_{1 \leq i \leq 4} r_i \left(H_i p_{(x, y)}^{n(x, y)} \right) = \hat{r} \left(p_{(x, y)}^{n(x, y)} \right) \\ g(T(x, y)) &= \frac{1}{4} \sum_{1 \leq i \leq 4} g_i \left(H_i p_{(x, y)}^{n(x, y)} \right) = \hat{g} \left(p_{(x, y)}^{n(x, y)} \right) \\ b(T(x, y)) &= \frac{1}{4} \sum_{1 \leq i \leq 4} b_i \left(H_i p_{(x, y)}^{n(x, y)} \right) = \hat{b} \left(p_{(x, y)}^{n(x, y)} \right) \end{aligned} \quad (4.9)$$

(4.9) 式の結果は、3.3 で述べた、color-consistency の評価の過程で、既に得られている ((4.4) 式参照) ことに注意されたい。見えの視点依存性を保つ観点からは、上記のような単純平均よりも、むしろ対象とする光線との近接度によって、個々の reference view に適切な重み付けをするほうが望ましい [20] が、提案手法では処理の効率を優先する。なお、(4.9) 式を採用した場合でも、最終的な光線の色への寄与を、近傍の 4 つの reference view のみに限定しているため、ある程度の視点依存性は保たれる。

4.3.7 擬似コード

与えられた target view に対して画像合成を行う関数 ViewSynthesis() の擬似コードを図 4.3 に示す。引数 H_t は、target view の平面射影を表す斉次行列である。

関数の内部は、主に、(Step A) 初期化、(Step B) 各レイヤに対する繰り返し処理、(Step C) 画像の表示の 3 つの部分に分けられる。(Step A) では、 $E_{\min}(x, y)$ を初期化する。 $E_{\min}(x, y)$ は、各画素に対して、color-consistency の最小評価値を保持するのに用いるため、初期化の際には、十分に大きな値をセットする。(Step B) は、各レイヤに対する繰り返し処理である。まず、(4.5) 式

にしたがって、各画素に対する color-consistency の評価値を計算する．この計算の途上で得られる、RGB 各成分の平均値は、後に画素の色付けに用いるため、 $tmp(x, y)$ に蓄えておく．各画素に対する評価値は、 $E(x, y)$ に格納される．次に、 $E(x, y)$ を、(4.6) 式にしたがって平滑化する．最後に、各画素ごとに、 $E(x, y)$ と $E_{\min}(x, y)$ の大小を比較し、 $E(x, y)$ のほうが小さい場合は、最小値の置き換え、および画素 $T(x, y)$ への色付けを行う．すべてのレイヤに対して上記の処理を行うことにより、最終的に、 $T(x, y)$ は、(4.9) 式に従って色づけされることになる．最後に (Step C) で、 $T(x, y)$ を画面上に表示する．

全体の処理時間の中で、大半を占めるのは (Step B) である．合成画像の画素数を $X \times Y$ 、奥行きレイヤの枚数を N とすると、この部分の計算量のオーダーは、 $O(XYN)$ である．

4.4 実験

4.4.1 実装

入力画像 (reference view) として、「筑波大学多視点画像データベース」より、81 視点 (縦 $9 \times$ 横 9 のアレイ) で撮影された多視点画像群 (図 4.4) を利用した．各画像は、24 bit のカラー画像で、サイズは 256×192 画素とした．

実装環境は、Linux ベースの汎用の PC (Pentium 4 3.2 GHz, メインメモリ 2.0 Gbyte) 1 台である．グラフィックスカードには、NVIDIA GeForce FX 5800 のグラフィックスプロセッサ (GPU) と 128MByte のビデオメモリが搭載されている．ソフトウェアは、C 言語と OpenGL (一部は、ARB fragment program[21] を用いた) で記述し、gcc でコンパイルした．

Li ら [11] が示したように、(4.5) 式の計算は、マルチテクスチャ機能と fragment program の組み合わせによって、GPU 上で高速に実現できる．しかし、(4.6) 式の演算については、GPU 上で高速に実現するのは困難であった⁴．そこで、筆者らは、CPU 上での計算を併用して実装を行うことにした．

図 4.3 の (Step B) において、各レイヤに対する処理は、下記の順に行われる．

- (B.1): GPU 上 (OpenGL と fragment program) で、 $E(x, y)$, $tmp(x, y)$ を計算する．
- (B.2): $E(x, y)$, $tmp(x, y)$ をビデオメモリからメインメモリに転送する．
- (B.3): CPU 上で、 $E(x, y)$ 平滑化を行う．(4.7) 式を用いることにより、高速化される．
- (B.4): CPU 上で、 $E(x, y)$ と $E_{\min}(x, y)$ の大小比較、および $T(x, y)$ の色づけを行う．

4.4.2 処理時間

表 4.2 に、レイヤの数 (N) を変化させ、画像合成に要する処理時間を測定した結果 (100 回描画した平均値) を示す．(4.6) 式では、 $W = 3$ とした．画像を合成する視点位置に応じて、処理

⁴OpenGL1.2 以降では、GL_ARB_imaging 拡張仕様によって、畳み込み演算の機構がサポートされており、直接的に (4.6) 式の演算を実現できる．しかし、筆者らの環境では、この方法での平滑化処理は、非常に遅い (提案する実装の平滑化と比較して、100 倍以上の時間がかかる)．また、ARB fragment program には、畳み込み演算を行う命令が存在しない．

```

ColorImage  $T(x, y)$ ; /* target image */
ColorImage  $tmp(x, y)$  /* temporal image for each layer */
MonoImage  $E(x, y)$ ; /* evaluation image for each layer */
MonoImage  $E_{\min}(x, y)$ ; /* minimum of  $E(x, y)$  */

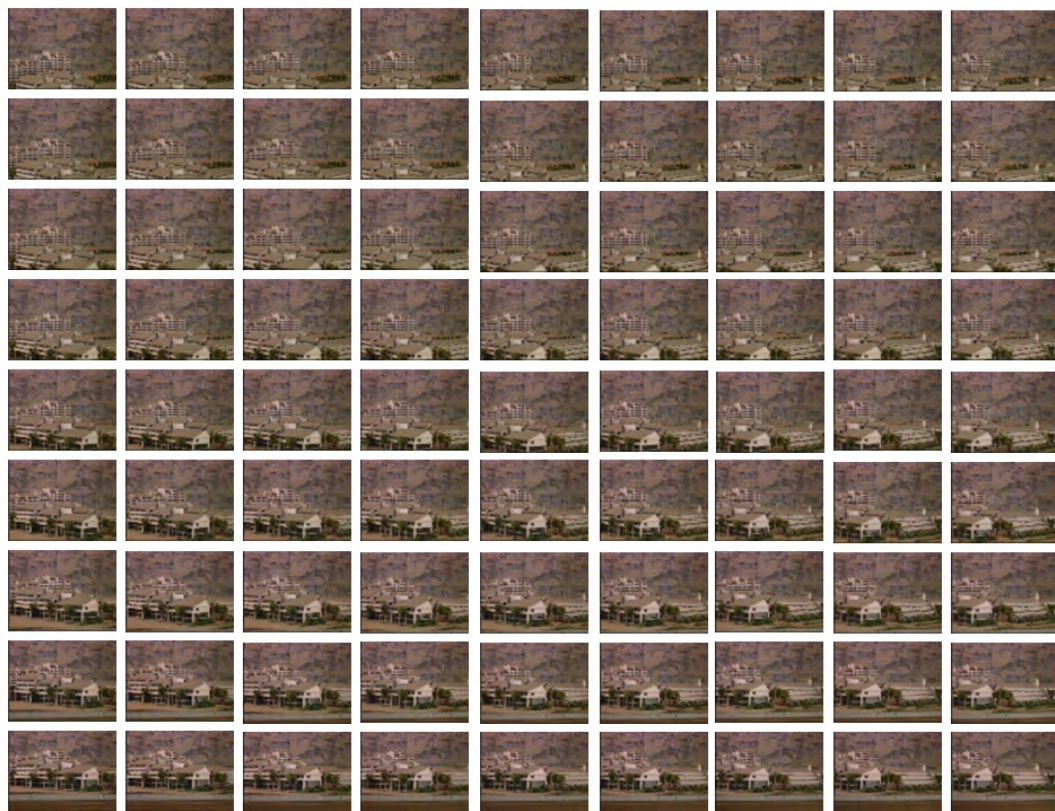
ViewSynthesis(target view  $H_t$ ){
  // initialization (Step A)
  Initialize  $E_{\min}(x, y)$ ;
  // iteration for each layer (Step B)
  for( $n = 1$ :  $n \leq N$ ;  $n++$  ){
    for(each  $(x, y)$ ){ /* consistency calculation (Eq.(4.5)) */
      calculate  $p_{(x,y)}^n$  for  $H_t$ ;
      load  $H_1, H_2, H_3, H_4$  for  $p_{(x,y)}^n$ ;
       $tmp(x, y) = \{ \hat{r}(p_{(x,y)}^n), \hat{g}(p_{(x,y)}^n), \hat{b}(p_{(x,y)}^n) \}$ ;
       $E(x, y) = \sigma_r^2(p_{(x,y)}^n) + \sigma_g^2(p_{(x,y)}^n) + \sigma_b^2(p_{(x,y)}^n)$ ;
    }
    for(each  $(x, y)$ ){ /* smoothing (Eq.(4.6)) */
       $E(x, y) = \hat{E}(x, y)$ ;
    }
    for(each  $(x, y)$ ){ /* pixel coloring (Eq.(4.8),(4.9)) */
      if( $E(x, y) < E_{\min}(x, y)$ ){
         $E_{\min}(x, y) = E(x, y)$ ;
         $T(x, y) = tmp(x, y)$ ;
      }
    }
  }
  // displaying (Step C)
  display  $T(x, y)$ ;
}

```

図 4.3: 擬似コード

時間に3–4%程度のばらつきがあることが観察されたが、表 4.2 には、最も時間がかかる視点位置に相当する数値を示した。target view のサイズは、 256×256 画素と、 512×512 画素の二通りである。いずれの画素数においても、処理時間はおおよそレイヤの枚数 (N) に比例しており、この点では理論に一致する結果が得られていることがわかる。しかし、レイヤの枚数を同一として、 256×256 画素の場合と、 512×512 画素の場合を比較すると、画素数の比は $1 : 4$ であるにもかかわらず、計算時間の比はおおよそ $1 : 3$ となっている。処理時間の内訳を詳細に調べることで、この理由について考察する。

表 4.3 に、(Step B) を構成する処理に関して、所要時間の内訳 (1000 回処理した平均値) を計



(a) city



(b) plants

図 4.4: 入力画像データ (「筑波大学多視点画像データベース」より)。

表 4.2: 画像合成に要する総処理時間 (ms): 81 枚の入力画像を使用

	$N=5$	$N=10$	$N=15$
256×256	42.7 (23.4 fps)	84.7 (11.8 fps)	126.5 (7.9 fps)
512×512	123.8 (8.1 fps)	244.4 (4.1 fps)	365.4 (2.7 fps)

表 4.3: 各レイヤに対する処理時間の内訳 (ms)

	(B.1)	(B.2)	(B.3)	(B.4)	total
256×256	2.69	3.13	2.09	0.49	8.40
512×512	2.88	11.31	8.69	1.30	24.18

表 4.4: 画像合成に要する総処理時間 (ms): 25 枚の入力画像を使用

	$N=5$	$N=10$	$N=15$
256×256	42.3 (23.6 fps)	82.4 (12.1 fps)	122.6 (8.2 fps)
512×512	124.2 (8.1 fps)	244.0 (4.1 fps)	363.0 (2.8 fps)

測した結果を示す (B.1) の処理時間は、画素数に関わらず、ほとんど変化しないことが分かる。これは、この部分が GPU 上で実装されており、内部で自動的に並列化/パイプライン化が行われているためと考えられる。この部分の処理が、全体の処理時間が画素数に比例しないことの主要因である。また、(B.2) のデータ転送が、4 つの処理の中で最大の割合を占めていることもわかる。全体のスループットを向上させるためには、並列処理などにより、このデータ転送の待ち時間を活用することが有効であると考えられる。特に、本章の実装においては、(B1) は GPU 上で実行し、(B3)、(B4) は CPU 上で実行するため、各レイヤに対する処理をパイプライン化することによって、スループットの向上が見込まれる。これらの点については、今後の課題とする。

最後に、入力カメラの数と処理時間の関係について調べた。前章で述べたように、提案手法の理論上の計算コストは、ボクセル数 (レイヤ数 \times 合成画像の画素数) に依存するが、カメラ数には依存しない。この点を確認するため、81 視点の入力画像のうち、25 視点分のみを用いて画像合成を行い、処理時間を計測した。具体的には、元の 9×9 カメラアレイのうち、奇数行・列のカメラ画像 25 枚 (5×5) のみを用いた。結果を表 4.4 に示す。表 4.2 と表 4.4 を比較すると、同一の解像度およびレイヤ数においては、81 視点の場合と 25 視点の場合とで、計算時間にほとんど差がないことが確認できる。これは理論を支持する結果である。なお、ボクセル数一定ではなく、品質を一定に保つ条件下では結論が異なる。25 視点分のみを用いる場合、カメラの間隔が元の 2 倍となるため、理論上、レイヤ数も 2 倍にする必要がある (3.2 参照)。したがって、この場合には、81 視点すべてを用いる場合と比較して、約 2 倍の処理時間を要する。

4.4.3 奥行き推定および画像合成の結果

図 4.5 は、街の模型を被写体とした多視点画像群 City を用いた、奥行き推定、および画像合成の結果である。target view のサイズは 256×256 画素、奥行きレイヤの数 $N=10$ 、(4.6) 式では $W=3$

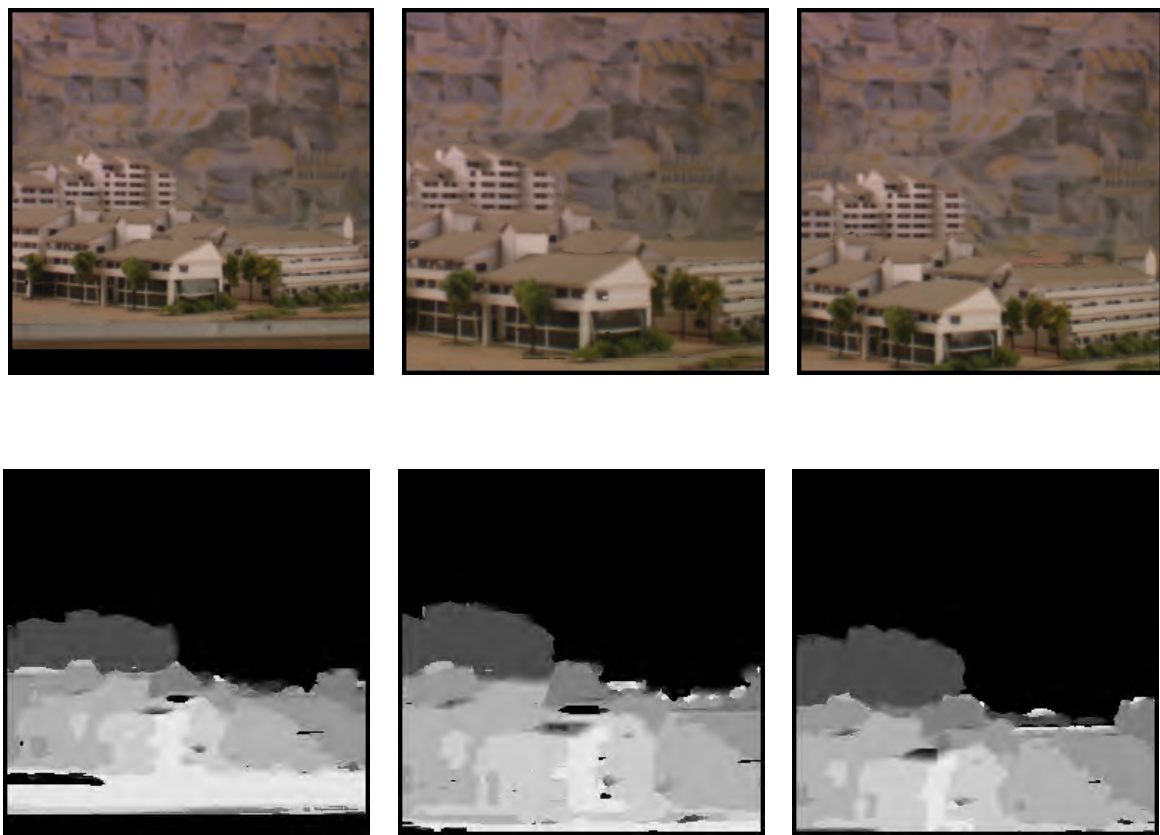


図 4.5: City 画像群を用いた合成画像（上段）と対応する奥行きマップ（下段）

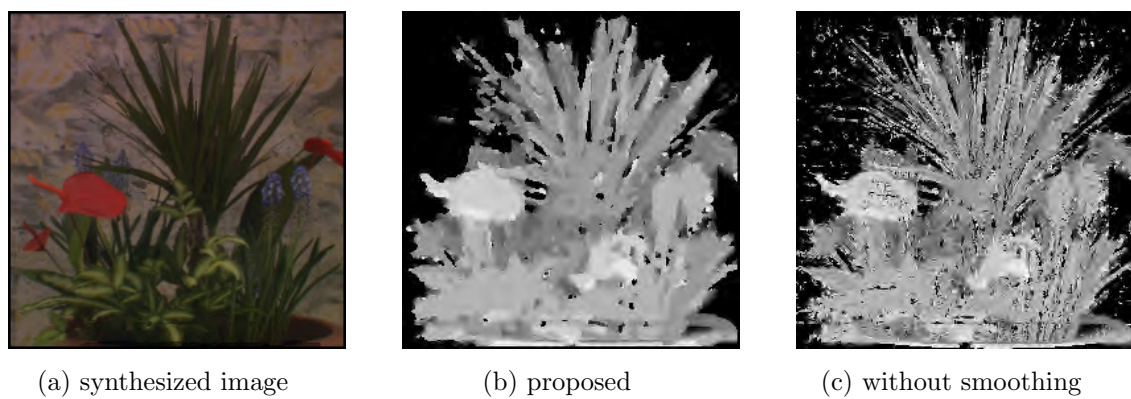


図 4.6: Plants 画像群を用いた合成画像と奥行きマップ

としている．与えられた視点位置に応じて，適切な奥行きマップが推定され，実写画像と見違えるような鮮明な画像が合成できていることがわかる．

図 4.6 は，植物を被写体とした多視点画像群 Plants を用いた結果である．図 4.6(a) および (b) では， $W=2$ としている．比較のため，(4.6) 式の平滑化を行わない場合（図 4.6(c)）の奥行きマップも併せて掲載する．図 4.6(c) から，平滑化を行わない場合，奥行き推定が不安定になり，奥行きマップにノイズが発生することがわかる．これらのノイズにより，合成画像上には，本来存在しない粒子状の浮遊物が多数発生するため，視覚的な品質が損なわれる．

提案手法では，一般の photo-hull の構築とは異なり，visibility のチェックを行っていない．したがって，奥行きが不連続になる部分（手前のオブジェクトと，遠方のオブジェクトの境界部分）では，オクルージョンの影響のため，正確な推定が行えない．推定された奥行きマップを見ると，手前のオブジェクトが「太る」傾向があることが見受けられる．提案手法に対して，何らかの形で visibility のチェックを導入したり，平滑化処理のブロックサイズを適応的に変化させる機能を加えることにより，このような細部の推定を改善できる可能性がある．その際には，処理時間と最終的な合成画像の品質とのトレードオフを考慮する必要がある．

4.5 むすび

本章では，2 次元カメラアレイを用いて取得した多視点画像をもとに，奥行きレイヤモデルを用いて自由視点画像合成を行うことを想定した，奥行き推定の手法について検討した．オフライン計算を行わず，動的なシーンを扱うことを念頭に置き，所望の視点位置での画像合成に必要な情報のみを含む，視点依存奥行きマップを高速に推定する手法を提案した．本手法の特徴は，合成画像の画素単位の奥行きを，高い精度で推定することにある．また，グラフィックスハードウェアのアクセラレーションを用いた実装により，実時間で，実写画像に遜色のない，高品質な自由視点画像を合成できることを確認した．

なお，本章で実験に用いた素材（「筑波大学多視点画像データベース」）は，1 台のカメラを用いて，アームロボットでその位置を制御しながら撮影されたものであり，理想的な条件に近い多視点画像であると考えられる．実際のカメラアレイを構築した場合には，キャリブレーションの誤差や伝送路での画像の劣化など，理想的な条件とは異なる要因が含まれ得る．今後は，新たなカメラアレイの構築や動的なシーンへの適用を視野にいれ，これらの誤差が奥行き推定や合成の品質に与える影響について，理論と実験の両面から検討を行う予定である．

参考文献

- [1] T. Kanade, P.J. Narayanan, and P.W. Rander, “Virtualized Reality: Concepts and Early Results,” IEEE Workshop on the Representation of Visual Scenes, pp. 69–76 (1995).
- [2] W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan, “Image-Based Visual Hulls,” Proc. ACM SIGGRAPH 2000, pp. 369–374 (2000).

- [3] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video-Based Modeling and Rendering of 3D Scenes,” *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, pp. 66–73 (2002).
- [4] N. Inamoto and H. Saito, “Fly Through View Video Generation of Soccer Scenes,” *Intl. Workshop on Entertainment Computing, Workshop Note*, pp. 94–101 (2002).
- [5] T. Matsuyama and T. Takai, “Generation, Visualization, and Editing of 3D Video,” *Intl. Symposium on 3D Data Processing, Visualization and Transmission*, pp. 234–245 (2002).
- [6] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera,” *Proc. 13th Eurographics Workshop on Rendering*, pp. 77–86 (2002).
- [7] R. Yang, G. Welch, and G. Bishop, “Real-Time Consensus-Based Scene Reconstruction Using Commodity Graphics Hardware,” *Proc. Pacific Graphics*, pp. 225–235 (2002).
- [8] G.G. Slabaugh, R.W. Schafer, and M.C. Hans, “Image Based Photo Hulls for Fast and Photo-Realistic New View Synthesis,” *Real-Time Imaging*, Vol. 9, No. 5, pp. 347–360 (2003).
- [9] I. Kitahara and Y. Ohta, “Scalable 3D Representation for 3D Video in a Large-Scale Space,” *Presence*, Vol. 13, No. 2, pp. 164–177 (2004).
- [10] C. Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” *Proc. Eurographics Symposium on Rendering*, pp. 243–254 (2004).
- [11] M. Li, M. Magnor, and H.-P. Seidel, “Hardware-Accelerated Rendering of Photo Hulls,” *Proc. Eurographics 2004*, Vol. 23, No. 3, pp. 635–642 (2004).
- [12] J.-X. Chai, X. Tong, S.-C. Chany, and H.-Y. Shum, “Plenoptic Sampling,” *Proc. ACM SIGGRAPH 2000*, pp. 307–318 (2000).
- [13] J. Shade, S.J. Gortler, L.-W. He, and R. Szeliski, “Layered Depth Images,” *Proc. ACM SIGGRAPH 98*, pp. 231–242 (1998).
- [14] A. Isaksen, L. McMillan, and S.J. Gortler, “Dynamically Reparameterized Light Fields,” *Technical Report MIT-LCS-TR-778* (1999).
- [15] X. Tong, J. Chai, and H.-Y. Shum, “Layered Lumigraph with LOD Control,” *The Journal of Visualization and Computer Animation*, Vol. 13, No. 4, pp. 249–261 (2002).
- [16] S.M. Seitz and C.R. Dyer, “Photorealistic Scene Reconstruction by Voxel Coloring,” *Proc. IEEE CVPR*, pp. 1067–1073 (1997).
- [17] K.N. Kutulakos and S.M. Seitz, “A Theory of Shape by Space Carving,” *Intl. Journal of Computer Vision*, Vol. 38, No. 3, pp. 199–218 (2000).

- [18] U.R. Dhond and J.K. Aggarwal, “Structure from Stereo – A Review,” IEEE Trans. Systems, Man, and Cybernetics, Vol. 19, No. 6, pp. 1489–1510 (1989).
- [19] A. Laurentini, “The Visual Hull Concept for Silhouette-Based Image Understanding,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 16, No. 2, pp. 150–162 (1994)
- [20] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, “Unstructured Lumigraph Rendering,” Proc. ACM SIGGRAPH 2001, pp. 425–432 (2001).
- [21] ARB_fragment_program, http://oss.sgi.com/projects/ogl-sample/registry/ARB/fragment_program.txt (2003).

第5章

合焦判定に基づく 全焦点自由視点画像合成 — 空間領域における検討

5.1 はじめに

本論文の目的は、「多視点画像を入力として、高品質な自由視点画像を、実時間で合成すること」である。この技術により、三次元動空間を対象とした実時間自由視点画像合成 [1, 2, 3, 4, 5, 6, 7, 8, 9] が可能となり、遠隔地通信やバーチャルリアリティ分野での応用が期待される。本論文の第4章では、上記の目的のために、奥行き推定処理をいかに最適化するか、という問題について論じた。

第4章で提案した手法の原理は、入力画像の画素値の相関を考慮することで、合成画像の各画素に対する最適な奥行きを探索する、というものである。これは、伝統的なコンピュータビジョン手法である多眼ステレオ法の枠組みの中で、自由画像合成に特化した最適化を行ったことに相当する。一方、本章では、奥行き探索の新しい方法論として、“合焦判定”と呼ばれる手法を提案する。この手法の特徴は、レイヤモデルを構成する個々の平面に対する合成画像を分析し、“合焦”領域を検出することで、各画素に最適な奥行きを与える点にある。さらに本章では、提案手法を1台のPC上で効率的に実装し、インタラクティブレートに近い速度でのレンダリングが可能であることを確認する。

本章においても、第4章と同様に、二次元アレイ状に構成された多眼カメラによる画像群を入力とし、奥行きレイヤモデルを用いて画像を合成する。この入力 *light field* データ [10, 11] と呼ばれる。もし、カメラが十分に密に配置されていれば、自由視点画像合成を行う際に、被写体の幾何学構造に関する知識は一切不要である。しかしながら、必要とされる入力画像の数を現実的な値に抑えるためには、画像に加えて、被写体空間の構造モデルを用いることが必要とされる。必要とされる入力画像の数と、必要とされる幾何モデルの詳細度との間にはトレードオフ関係があることが知られており、文献 [12] では、この関係は最小サンプリング曲線 (minimum sampling curve) として定式化されている。奥行きレイヤモデルを用いて画像を合成する場合には、その画像の各画素がどのレイヤに属しているのかを判定する必要がある。この判定を、本論文では、奥行き割当てと呼ぶことにする。

奥行きレイヤによって被写体空間の幾何学的形状を表現する手法は、IBR においてよく用いられてきた。初期のいくつかの研究 [13, 14] では、被写体の形状情報が既知であることを前提とし、その情報を用いて奥行き割当てを行っていた。被写体の形状情報を得るためには、通常、前処理 (オフライン計算) が必要となるため、これらの手法は、動的なシーンを扱うには不向きであった。さらに近年では、計算機の進歩により、形状復元手法を簡略化・効率化して実時間処理に組込む、というアプローチ (image-based modeling and rendering と呼ばれる) も現実的なものになりつつある。このようなアプローチの元に、複数の研究者により、興味深い結果が多数報告されており [3, 5, 6, 7, 8, 9]、本論文の第4章の提案手法もこの範疇に属する。一方、本章で導入する手法は、合成画像の分析に基づく特有の奥行き割当てアルゴリズムを採用しており、処理の枠組みにおいて、従来の image-based modeling and rendering とは一線を画するものである。



図 5.1: Light field rendering における焦点ぼけ (文献 [14] より引用) .

5.2 背景

本章で導入する奥行き割り当ての手法は，“合焦判定”と呼ばれる．本節では，この言葉の意味を明確にするため，light field rendering とその“焦点ぼけ”現象について，および light field に奥行きレイヤモデルが導入されていった経緯について概説する．

1990 年代の初頭以降，三次元視覚情報の最小単位として，「光線」が注目されるようになった [16, 17, 18]．特に，コンピュータグラフィックスの分野において，Levoy と Hanrahan [10] が提案した light field rendering 法は，後の研究に与えた影響が大きい．この手法は，密に取得された多視点画像を入力とし，写実性の高い自由視点画像を合成するものである．入力多視点画像は，四つの変数 (s, t, u, v) を用いて光線の集合としてパラメータ化され，light field データを構成する．もともと，この手法は，入力多視点画像が十分に高い密度で取得されることを前提としていた．しかしながら，文献 [12] で示されたように，その前提を満たすためには，極めて小さなカメラ間隔で，膨大な数の画像を取得する必要がある．

Isaksen ら [14, 19] は，画像合成の際に任意に配置できる focal surface を導入することにより，light field rendering 法を拡張した．focal surface は任意の形状を取ることができ，シーンの構造を近似するモデルに相当する．また彼らは，light field データのサンプリング密度が不十分な場

合, focal surface の位置に応じて, “焦点ぼけ” 現象が生じることを示した. 例として, Isaksen らの手法において一枚の平面 (平面なので focal plane と呼ぶ) をモデルとして用いて画像を合成した場合を考える. 図 5.1 では, focal plane の位置を順に手前から奥に移動した様子を示している. focal plane 付近にある物体は鮮明に合成される (“焦点” が合う) が, focal plane から離れた物体には, ぼけや二重像が生じる (“焦点” から外れる) ことが分かる. “合焦” 距離は, focal plane の奥行きによって制御される. これらの合成画像においては, いずれも同一の入力多視点画像 (鮮明な画像) が用いられている点に注意されたい. ここで述べる “焦点ぼけ” は, 画像合成の過程に起因する現象である (物理カメラの焦点ぼけとの関係については, 文献 [20] の考察が興味深い).

さらに Isaksen らは, 下記の手順により, サンプリング密度が不十分な light field データから, 全体が鮮明な画像 (“全焦点” 画像) を合成する手法を示した.

1. 与えられた視点位置において, focal plane の奥行きを変えながら, 複数枚の画像を合成する.
 2. それらの合成画像の “合焦” 領域を検出し, 一枚の画像に統合することで, “全焦点” 画像を得る.
1. は, 対象シーンにレイヤ群を配置することに相当する. 2. は, 各画素に対して最適な奥行きを与える, 奥行き割り当て処理に相当する.

このアプローチにおいて鍵となるのは, 2. において, どのように “合焦” 領域を見つけるか, という点である. この処理を行うにあたり, Isaksen らは, オフライン処理により, あらかじめ三次元モデルが構築されていることを前提とした. 彼らの用いた三次元モデルとは, レイヤ群の各点において, 被写体が存在する確からしさが与えられたものであった. この手法は, 視覚的に良好な画像を合成する点で有効であることは示されたが, オフライン計算を必要とするため, 動的なシーンを扱うには不向きであった. 一方, 本章における提案手法は, ランタイムに動作するものであり, 1. で合成された画像を直接分析することにより, その “合焦” 領域を検出するものである.

物理的なカメラによって撮影される一般の画像の場合, 合焦領域の検出には, 高周波成分を用いることが一般的である [21]. この合焦尺度は, 焦点ぼけが, 低域通過フィルタとしてモデル化されることを前提としている. しかし, light field rendering における “合焦” 現象は, 一般のカメラのものとは異なっている. Light field rendering による合成画像では, “焦点” から外れる領域には, ぼけだけではなく, 高周波を含む二重像が発生する. 特に, 視覚的品質の点で問題になるのは, 後者の二重像である. したがって, 本章における提案手法では, 一般の画像の場合とは異なる “合焦” 尺度を定義する.

5.3 提案アルゴリズム

図 5.2 に示すように, 入力多視点画像は, ほぼ 1 枚の平面上に配置された多数のカメラによって取得されるものとする. カメラが配置される平面を $Z = 0$ とする. カメラ群はおおよそ 2 次元アレイ状に配置されることを前提とするが, 各カメラの位置・姿勢には厳密な制約を設けない. た

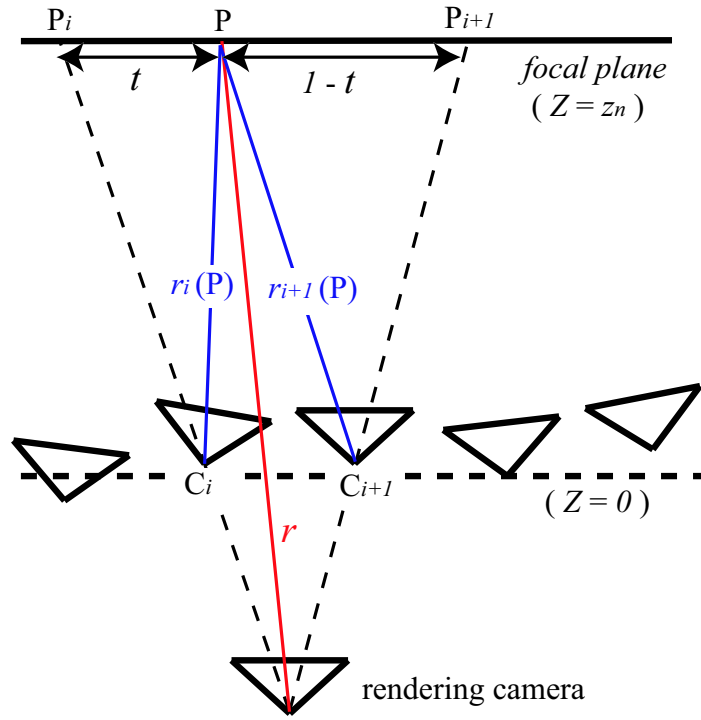


図 5.2: 基本設定

だし，各カメラは事前にキャリブレーションされている（内部および外部パラメータが計測されている）ことを前提とする．提案手法の目的は，図中の rendering camera（位置，姿勢や任意に制御できる）において画像を合成することである．なお，本節では，簡単のためカメラアレイを一次元として議論するが，本来対象とする二次元アレイへの拡張は自明であるため割愛する．

画像合成の際には，奥行きレイヤモデルを用いる．レイヤの枚数を N ， n 番目のレイヤの奥行き（カメラを配置した面からの距離）を $Z = z_n$ で表す．シーンの幾何学的構造は未知であるものの，奥行きの最大値 Z_{\max} と最小値 Z_{\min} は既知であるとする．文献 [12] に従い，視差空間 ($1/Z$) を等間隔に分割するように， N 枚のレイヤを配置することにする．

$$\frac{1}{z_n} = \frac{1}{Z_{\max}} + \frac{n-1/2}{N} \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) \quad (5.1)$$

提案手法において，ある視点位置に対する合成処理は，図 5.4 のような流れになる．これは，次のような三段階に分けて考えることができる．

Step. 1 各レイヤに対して画像を合成する． n 番目のレイヤに対する合成画像を $I_n(x, y)$ とする．これらの画像では，配置されたレイヤ付近の被写体のみが鮮明に（“焦点” が合った状態で）合成される．

Step. 2 $I_n(x, y)$ ($n = 1, 2, \dots, N$) の各画素について，“合焦”の度合いの評価値 $f_n(x, y)$ を計算する．

Step. 3 上記の評価値を用いて, $I_n(x, y)$ ($n = 1, 2, \dots, N$) から, 最も“焦点”が合っているとされる部分を抽出し, 最終的な1枚の画像 $I(x, y)$ に統合する.

ここで, (x, y) は画素位置を表す. 以下, 各段階の詳細な説明を行う.

5.3.1 各レイヤに対するレンダリング

図5.2を用いて, 一枚の focal plane を用いたレンダリング手法を説明する. ここで focal plane は, レイヤモデルを構成する各レイヤに対応する. 先ほども述べたように, 図中の rendering camera は, 合成画像を生成する視点位置を表す. C_i は i 番目の入力カメラを表し, rendering camera を中心に, C_i の中心を平面モデル (レイヤ) に投影した点を P_i とする. 画像合成は, rendering camera の投影中心に到達する個々の光線の色を求めることによって実現される.

図5.2において光線 r の色を求める場合について考える. 光線 r と平面モデルとの交点を P とする. ここで, 点 P は, 線分 $P_i P_{i+1}$ を $t : 1 - t$ に内分するものとする. 点 P を通り, 近傍のカメラ C_i, C_{i+1} に到達する光線を, それぞれ $r_i(P), r_{i+1}(P)$ とする. 光線 r の色 $c(r)$ は, 次式に従って求められる.

$$c(r) = (1 - t) \cdot c(r_i(P)) + t \cdot c(r_{i+1}(P)) \quad (5.2)$$

上記のような平面モデルと補間に基づく画像合成手法そのものは, よく知られた方法である [10, 19]. ここで重要なのは, 上記の手法で合成される画像では, 平面モデル付近の物体のみが鮮明に合成され (“焦点” が合う), 平面モデルから離れた物体にはぼけや二重像が発生する (“焦点” から外れる) ことである (図5.8 (a)–(c) 参照). 被写体空間全体を鮮明に合成するためには, 複数枚の平面モデル (奥行きレイヤ) を用いることが必要である.

5.3.2 “合焦尺度” の計算

次のステップは, 合成画像上で “焦点” が合っている部分を検出するための評価尺度を計算することである. この時点では, Step. 1 により, 各レイヤにつき一枚の画像が既に合成されている. 我々のアイデアは, 各レイヤに対してもう一枚の画像 (参照画像) を合成し, Step. 1 による画像と, 参照画像を比較することで, “焦点” が合っている領域を検出する, というものである.

例として, ある視点位置に対して, 図5.3に示した二つのモードによって画像を合成した場合を考える. 基本モードでは, すべての入力画像を合成に使用する. これは, Step. 1 における合成に対応する. 一方, 参照モードでは, カメラを一つおきにサブサンプルして, 選ばれたカメラのみを用いて画像を合成する. 図中では, 使用されないカメラを点線で示した. 参照モードによる画像は, Step. 2 における “合焦尺度” の算出のみに用いられる. どちらのモードでも, 各光線の色を求める際には, (5.2) 式を用いる. 図5.3では, 平面モデルは Object1 の位置に配置されているため, Object1 には “焦点” が合い, Object2 は “焦点” から外れる.

以下の議論では, 非拡散反射とオクルージョンが無視できると仮定する. “焦点” が合う部分から発している光線 (光線 r) の色は, どちらのモードでも同一の合成結果となる. これは, どちら

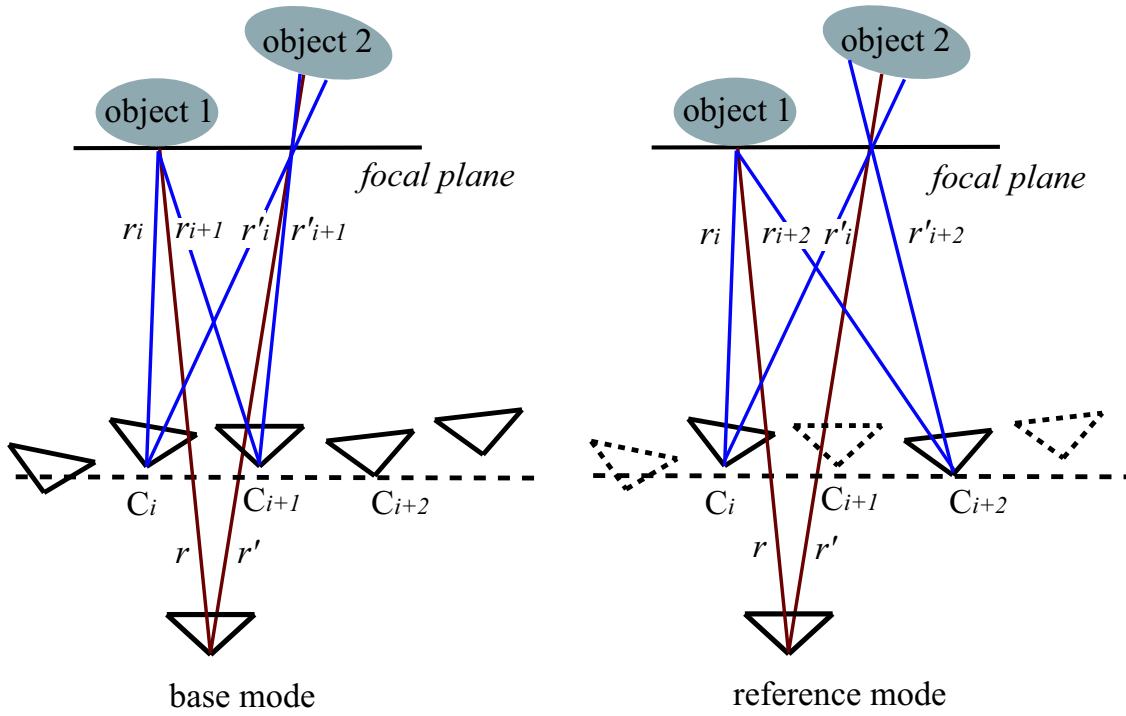


図 5.3: “合焦判定” のための二つの合成モード

のモードにおいても、被写体上の同一の点から発する光線が合成に用いられるためである（基本モードでは r_i, r_{i+1} が、参照モードでは r_i, r_{i+2} が合成に使用される）。一方、“焦点” から外れた部分から発する光線（光線 r' ）は、モードに応じて異なる色となり得る。これは、合成に用いられる各光線（ r'_i, r'_{i+1}, r'_{i+2} ）が、被写体表面上の異なる点から発するからである。したがって、二つのモードによって合成された画像を比較すると、“焦点” が合う部分は同一の合成結果となるが、“焦点” から外れた部分には違いが生じ得ることになる。

上記の議論に基づき、我々は、点 (x, y) 、奥行き $Z = z_n$ に対する“合焦尺度” $f_n(x, y)$ を次のように定義する。

$$sub_n(x, y) = |I_n(x, y) - R_n(x, y)| \quad (5.3)$$

$$f_n(x, y) = \sum_{-M \leq k \leq M} \sum_{-M \leq l \leq M} \frac{sub_n(x+k, y+l)}{(2M+1)^2}. \quad (5.4)$$

$I_n(x, y)$ および $R_n(x, y)$ は、 $Z = z_n$ に配置した平面モデルに基づいて、基本モードおよび参照モードで合成された画像を表す。(5.4) 式の平滑化処理は、“合焦判定” を安定化する働きをする（ M は正数）。ある画素 (x, y) について、 $f_n(x, y)$ が十分に小さな値をとる場合、その画素は、画像 $I_n(x, y)$ 上で“焦点” が合っていると判定される。

この“合焦”尺度は、多眼ステレオ法 [23] やボクセル復元法 [22] で用いられる奥行き推定の評価尺度と類似する。これらの手法では、複数の実カメラ画像において対応画素領域の一致度が評価される。これは、図 5.3 において、光線 r_i, r_{i+1}, r_{i+2} などの色を直接比較していることに相当する。一方で、提案手法では、光線 r_i, r_{i+1}, r_{i+2} の重み付け和により合成された光線同士の色

を比較している．原理的には，これらの尺度はすべて同一の仮定（空間中のある一点は，異なる方向から見ても，ほとんど同じ色に見える）に基づくものである．しかし，アルゴリズム上の相違は重要である．提案手法は自由視点画像合成を目的とするものであり，所望の視点位置における合成画像を直接分析する（(5.3)(5.4) 式）ことにより，各画素に対して，“合焦”の評価値を計算する点に特徴がある．

5.3.3 “全焦点”画像の合成

すべてのレイヤに対して“合焦”尺度 $f_n(x, y)$ が計算できれば，残りの処理は単純である．ある画素 (x, y) に対する最適な奥行きインデックス n_o は，“合焦尺度”の値の最小値探索によって決める．

$$n_o(x, y) = \arg \min_n (f_n(x, y)). \quad (5.5)$$

最後に，基本モードによる合成画像群 $I_n(x, y)$ ($n = 1, \dots, N$) を選択的に統合することにより，最終的な“全焦点”画像 $I(x, y)$ を得る．

$$I(x, y) = I_{n_o(x, y)}(x, y). \quad (5.6)$$

5.3.4 奥行き情報の平滑化

ここで述べる処理は，最終的な合成画像の品質を向上させるための，オプションである．(5.5) 式で与えられる奥行き情報では，しばしば，推定誤差によって，奥行きが周囲と不連続になる部分が生じる．特に，近くの被写体と遠くの被写体が重なり合っている場合に，その境界付近にノイズが生じやすいという現象が観察される¹．(5.6) 式によって得られる最終的な合成画像では，これらの部分に「ごみ」が浮いているように見える場合がある．特に，インタラクティブに視点位置を動かしたときに，これらの「ごみ」がちらつくように見えるため，視覚的な品質が損なわれる．

そのため，実時間性を損なわない範囲で，奥行き情報に対する平滑化を行うことを考える．このとき，ノイズ成分は抑制されることが望ましいが，奥行きの細かな変化も保持したい．これらの条件を満たす平滑化手法として，次のような手順を採用する．

まず，各点に対する“合焦”尺度 F の最小値を，その点の奥行きの信頼度の評価値 $R(x, y)$ と定義する． $R(x, y)$ が小さいほど，その点の奥行き情報の信頼性は高いと考える．

$$R(x, y) = \min_{0 \leq n \leq N-1} F^{(n)}(x, y) \quad (5.7)$$

次に，(5.5) 式で得られた $n(x, y)$ を単純にブロック平滑化したものを $n'(x, y)$ と表す． $n'(x, y)$ は， W を正の整数として次の式により与えられる．

$$n'(x, y) = \sum_{-W \leq i, j \leq W} \frac{n(x+i, y+j)}{(2W+1)^2}. \quad (5.8)$$

¹“合焦”判定法は，オクルージョンの影響が無視できることを前提にしているが，物体同士が重なり合う境界の部分では，その前提が成り立たない．

平滑化の処理は、(5.9) 式にしたがって行う。 R_{th} は信頼度評価値に対する閾値である。各画素について、信頼度評価が閾値以上の場合は、奥行きの値を、ブロック平滑化された値で置き換える。

$$n(x, y) = \begin{cases} n(x, y) & R(x, y) \leq R_{th} \\ n'(x, y) & \text{otherwise} \end{cases} \quad (5.9)$$

上記の処理では、“信頼度” の高い画素は、平滑化の影響を受けないため、奥行きの細かい変化が保存される。

5.4 実装と実験

5.4.1 実装

実装には、Pentium 4 3.2 GHz の CPU と 2.0 GB のメモリーを搭載した 1 台の PC を利用した。グラフィックスカードには、NVIDIA GeForce 5800 のグラフィックスプロセッサと 128 MB のビデオメモリーが搭載されている。ソフトウェアは C 言語と OpenGL で記述した。擬似コードは下記の通りである。本実装においては、5.3.4 で述べた奥行き情報の平滑化については省略する。

```
for (novel viewpoint){
    /* Stage A: synthesize N images by each mode */
    for (n := 1 → N){
        Synthesize  $I_n(x, y)$ ;
        Synthesize  $R_n(x, y)$ ;
    }
    /* Stage B: calculate focus measure values */
    for (n := 1 → N){
        Calculate  $sub_n(x, y)$ ; /* by Equation (5.3) */
        Calculate  $f_n(x, y)$ ; /* by Equation (5.4) */
    }
    /* Stage C: synthesize the final image */
    Calculate  $n_o(x, y)$ ; /* by Equation (5.5) */
    Calculate  $I(x, y)$ ; /* by Equation (5.6) */
}
```

ステージ A では、平面モデルを用いて、各モードに対して N 枚の画像を合成する。この部分については、マルチテクスチャリング機能を用いたグラフィックスハードウェア上での実装法 [19] を採用した。この処理では、各入力画像は、aperture テクスチャと呼ばれるグレースケールのテクスチャ画像によって変調され、射影テクスチャ座標を用いて平面モデルに投影される。aperture テクスチャは、各入力画像に対する重み付けの方法を決めるものである。本論文では (5.2) 式にしたがった重み付けがなされるように、このテクスチャを設計する。射影テクスチャ座標は、各

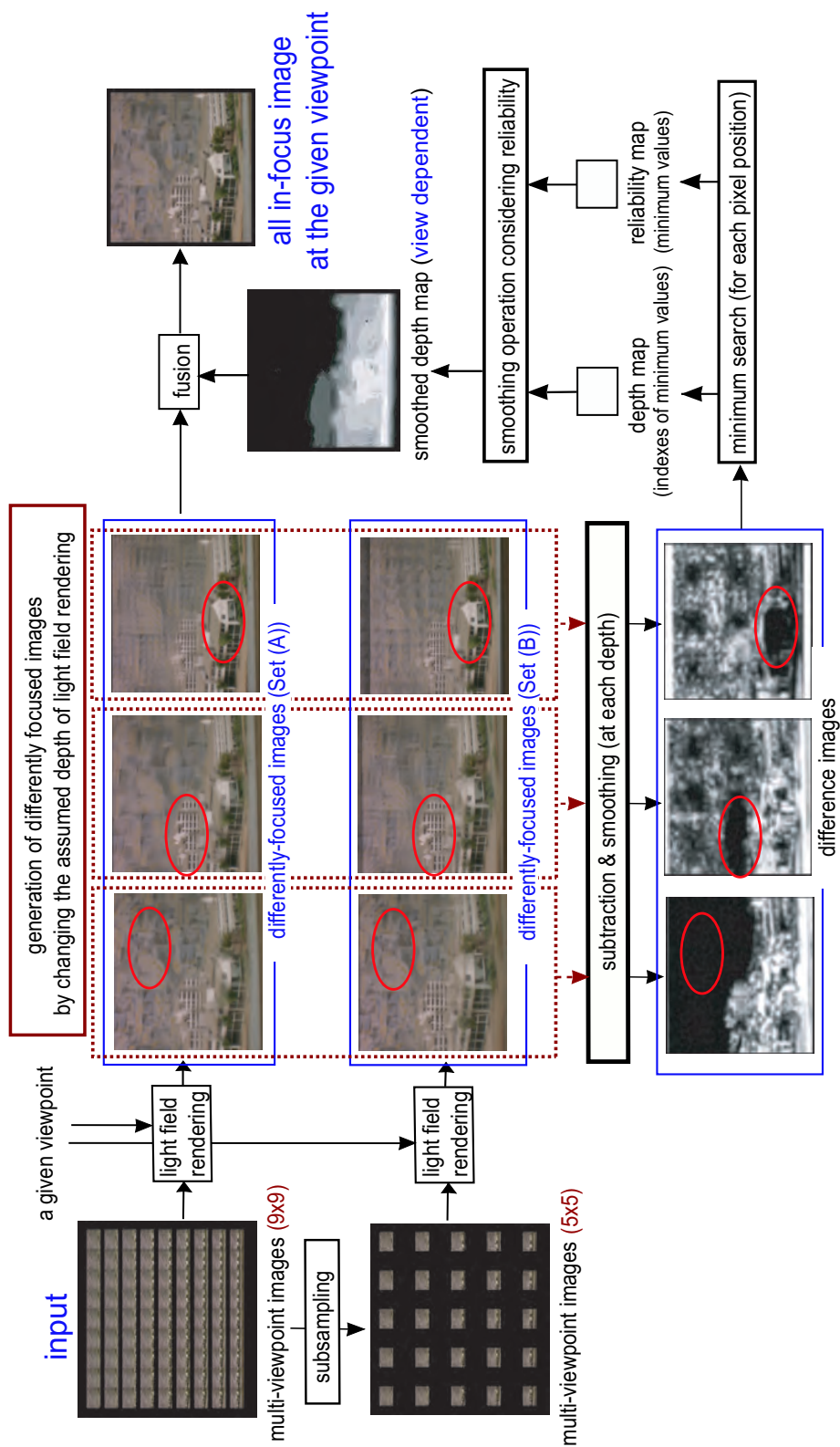


図 5.4: 提案手法のフローチャート.

カメラのキャリブレーション情報から導出されるものである．この座標を用いることにより，各カメラと平面モデルとの幾何学的な関係を正しく反映した投影処理（テクスチャマッピング）が実現される．上記の投影処理をすべての入力画像に対して行うことにより，一枚の合成画像が得られる．筆者らの環境では，上記の実装により，毎秒 1000 枚以上の画像を合成できる．合成された画像は，次の処理のため，メインメモリに転送される．

ステージ B では，すべての n に対して，“合焦尺度”を計算する．(2) 式 ($O(M^2)$) を直接計算する代わりに，次のような漸化式を用いることにより，計算オーダを $O(M)$ に下げることができる．

$$f_n(x, y) = f_n(x-1, y) + \sum_{-M \leq l \leq M} \frac{sub_n(x+M, y+l)}{(2M+1)^2} - \sum_{-M \leq l \leq M} \frac{sub_n(x-M-1, y+l)}{(2M+1)^2}. \quad (5.4')$$

筆者らの実装において，計算時間の大半は，ステージ A と B によって占められる．しかし，これらの処理は部分的に並列化できる．例えば， sub_1 と f_1 の計算は， I_1 と R_1 の計算が終わった時点で開始することができる．この性質を生かして計算時間を削減するため，筆者らはマルチスレッドを用いて実装を行った．

5.4.2 実験

正方格子カメラアレイ入力に対して

「筑波大学多視点画像データベース」に含まれる，サンタクロースの人形を被写体とした 81 視点の多視点画像（図 5.5）を入力として実験を行った．これらの画像は，一台のカメラを用いて， 9×9 の二次元格子点上で撮影されたものである．これらの画像群は，撮影条件が精密に制御されているため，等間隔でサンプリングされた light field データとみなすことができる．実際に複数のカメラを用いてアレイを構成する場合には，上記のような条件を整えることが極めて困難であることに注意されたい．

本実験においては，入力画像のサイズを 256×192 画素，合成画像を 320×240 画素とした．また，レイヤの枚数 (N) を 7，(5.4) 式における平滑化処理のパラメータ M を 3 とした．本実験においては，各入力カメラの撮像面がレイヤ群に正確に正対するため，射影テクスチャ座標は使用せず，通常のテクスチャマッピングを用いた実装を採用した．図 5.6 に，提案手法による合成画像を示す．比較のため，一枚のレイヤのみ ($Z = z_4$) を用いた合成画像も示す．提案法により，全体が鮮明で視覚的に良好な画像が合成できていることがわかる．ステージ A, B の並列化によりフレームレートは 13.9 fps となり，インタラクティブなレンダリングに十分な速度を達成できた．

非格子カメラアレイ入力に対して

入力として，カーネギーメロン大学先端マルチメディア研究室によって提供された多視点画像データ²を利用した（図 5.7 参照）．これらの画像は，キャリブレーション済みの 48 台のカメラ（縦 6 行 \times 横 8 列）によって撮影されたもので，サイズは 320×240 画素である．カメラ群はおお

²<http://amp.ece.cmu.edu/projects/MobileCamArray/>



図 5.5: 正方格子カメラアレイによる入力画像群 (「筑波大学多視点画像データベースより」).



(a) $z = z_4$

(b) proposed

図 5.6: 正方格子カメラアレイ画像を用いた実験結果

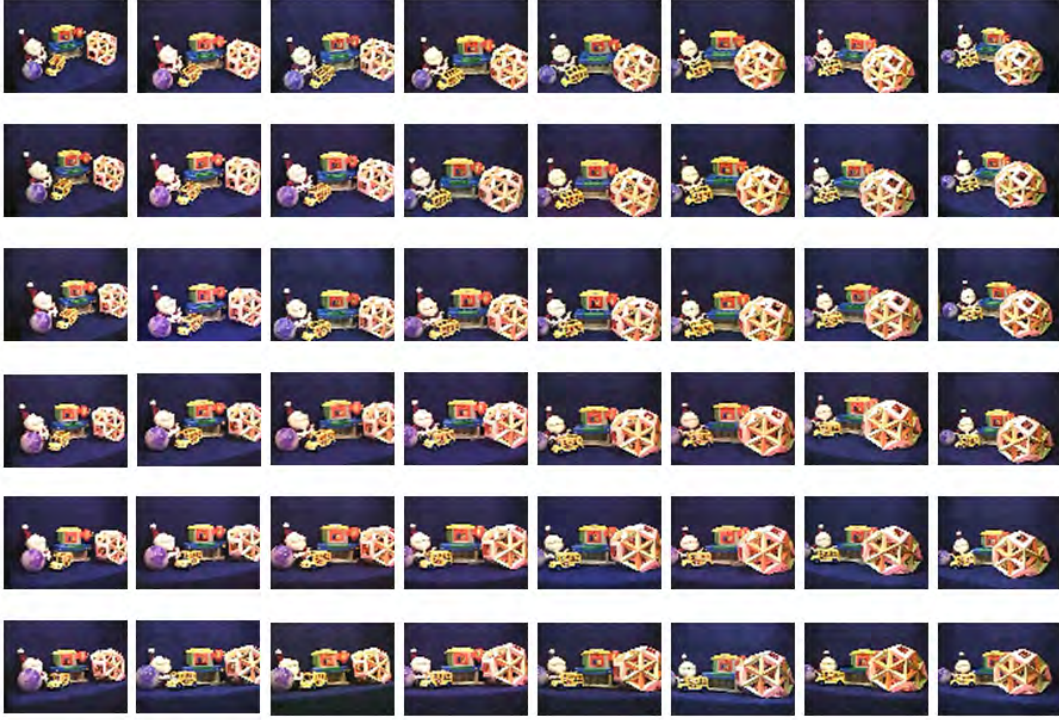


図 5.7: 入力多視点画像群（カーネギーメロン大先端マルチメディア研究室提供）

よそ平面上に配置されているが、カメラの間隔や向きは一定ではない。したがって、上記のデータは、非格子カメラ群入力の好例と考えられる。

本実験では、合成画像のサイズを 320×240 画素、奥行きレイヤの枚数 (N) を 7, (5.4) 式において $M = 5$ とした。さらに“合焦判定”を安定化するため、四つの参照モードを用いた。これらのモードは、入力カメラシステム上で、それぞれ（奇数/偶数）行と（奇数/偶数）列をスキップした場合に対応する。 sub_n を計算する際には、(5.3) 式の代わりに次の式を用いる。

$$sub_n(x, y) = \sum_j |I_n(x, y) - R_n^j(x, y)|. \quad (5.3')$$

ここで、 $R_n^j(x, y)$ は、 j 番目の参照モードにより、平面モデルを $Z = z_n$ に配置して合成した画像を表す。

図 5.8 は、ある視点位置における合成の過程を示したものである。(a), (b), (c) は、それぞれ、 $Z = z_2, z_5, z_7$ の位置に置かれた平面モデルに基づいて合成された画像である。ここで奥行きのインデックスは、奥から手前の順に付けられている。これらの画像には、“焦点”の影響が強く現れており、“焦点”から外れた部分には大きな劣化が生じていることがわかる。(d)–(f) は、(a)–(c) に対応する参照画像である。5.3 で述べたように、基本モード画像 ((a)–(c)) と参照画像 ((d)–(f)) において、“焦点”が合っている部分はほとんど同じように合成されており、逆に“焦点”から外

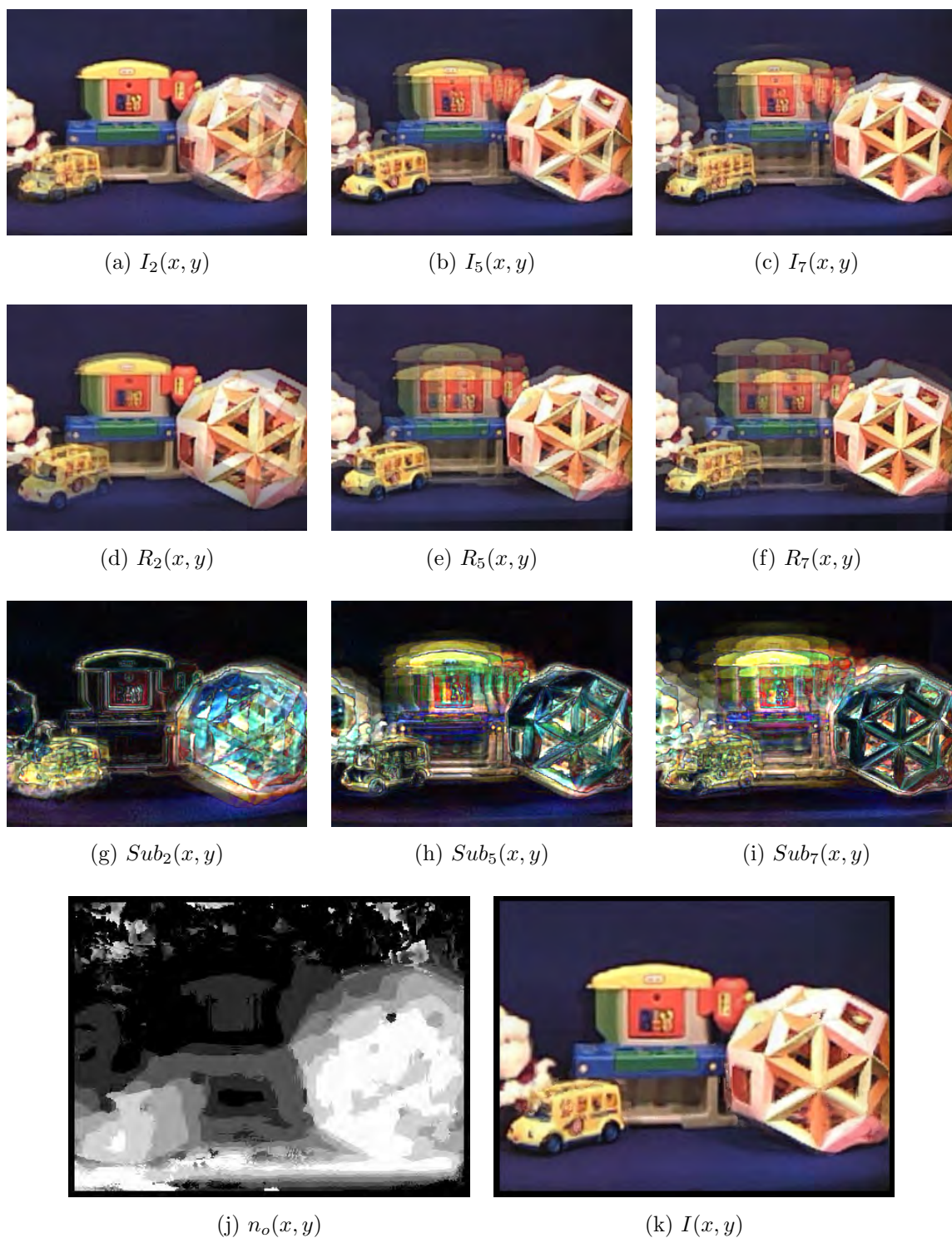


図 5.8: ある視点位置に対する合成過程 : (a)–(c) 平面モデルの奥行きを変化させながら合成した画像 , (d)–(f) 参照画像 , (g)–(h) 差分画像 , (j) 割り当てられた奥行き , および (h) 最終的な “全焦点” 画像 .

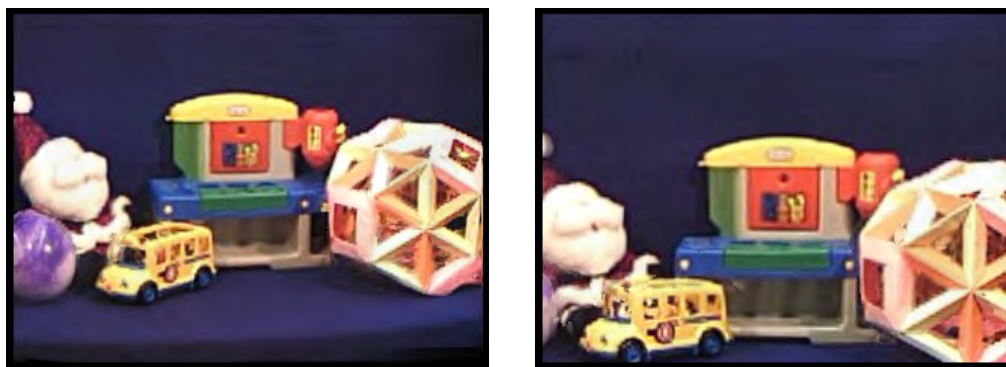


図 5.9: 他の視点に対する合成結果

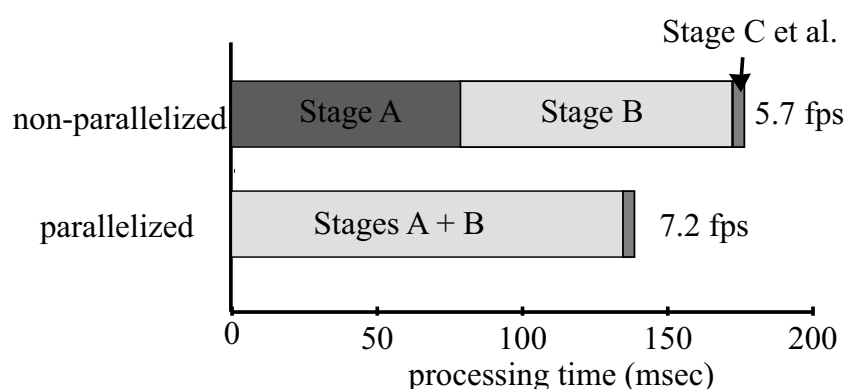


図 5.10: ステージごとの処理時間の内訳

れている部分には相違があることがわかる．したがって，提案手法では，(g)–(i) に示すような差分画像を“合焦判定”のために用いる．背景などのテクスチャがない部分は，どの奥行きにおいても，“焦点”が合っていると判定され得るが，これは最終的な合成画像の品質には影響しない．(j) に奥行きマップ（各画素に割り当てられた奥行きを明るさで示したもの），(k) に最終的な“全焦点”画像を示す．図 5.9 に示すように，他の任意の視点についても，高品質な“全焦点”画像を合成することが可能である．

図 5.10 には，各ステージごとの処理時間の内訳を示す．ステージ A と B を並列化した場合，並列化しなかった場合と比較して，全体の処理時間は約 20% 短縮された．今後一層の高速化が望まれるが，現時点でもインタラクティブレートに近い速度（7.2 fps）での処理を実現できている．正方格子カメラレイの場合と比較して，2 倍近い計算時間を要するのは，“合焦”判定の安定化のため，四つの参照モードを用い，さらに M の値をより大きくしているためである．これらのチューニングは，入力データの性質に依存するものである．

5.5 むすび

本章では，カメラレイによる多眼画像を入力として，奥行きレイヤモデルを用いて自由視点画像を合成する，新しい IBR 手法を提案した．各カメラが事前にキャリブレーションされていれ

ば、カメラの位置や向きを正確に合わせる必要はない。提案手法は、被写体の形状や奥行きがあらかじめ推定されていることを前提とせず、「合焦判定」という独自の手段により、実時間で画素単位の奥行き割当てを実現するものである。さらに、本手法の有効性を示すため、実装や実験の結果についても報告した。今後は、さらなる高速化の検討に加えて、動空間を対象とした実時間自由視点画像合成システム [3, 9, 4, 8] への適用についても検討する。

本章において提案した「合焦」判定法は、任意の視点位置における合成画像を直接分析することで、奥行きの確からしさを判定する新しいアプローチである。これは、自由視点画像合成という目的に対する直接的な手法であるだけでなく、「合成に基づく分析」という新しい方法論を導入した点で意義のあるものと言えよう。処理フローの観点から、従来のアプローチと提案手法の違いについて簡単にまとめておく。

図 5.11 に示すように、従来の image-based modeling and rendering においては、三次元形状推定 (modeling) と画像合成 (rendering) は直列的な関係にある。一般に、三次元形状 (3D geometrical model) は、レンダリングのパラメータとは無関係に求められるものである。したがって、静的な被写体を対象とする場合には、形状推定は事前にオフラインで行うことが一般的である。第 4 章では、全プロセスを実時間化するという目的のもとに、modeling を所望の視点位置における画像合成に必要十分な範囲に限定するアプローチを採用した。この場合、処理フローは図 5.12 のようになる。このアプローチでは、modeling と rendering を同時に行う場合もあり、必ずしも形状モデルが明示的に構築されるわけではない。しかし、奥行き (view-dependent depth information) が求まってから色が決まる、という順序においては図 5.11 と等価である。

提案手法の処理フローを図 5.13 に示す。このアプローチでは、まず、平面モデルを仮定して、仮の自由視点画像 (free-viewpoint images (interim)) を多数合成する。続いて、これらの画像を分析 (analysis) して奥行き情報 (view-dependent depth information) を推定し、さらにその結果を利用して最終的な「全焦点」画像を合成する (fusion)。一見すると、5.12 と比較して処理フローが複雑になっているが、本章の議論から明らかなように、提案手法は極めて単純に実装される。この新しい枠組みの可能性や理論的位置づけについては、第 6 章および第 8 章で、より発展的な議論を展開する。

参考文献

- [1] 苗村 健, 原島 博, “Video-Based Rendering の基礎検討”, 3 次元画像コンファレンス’98, 5-6, pp. 165–170 (1998).
- [2] Y. Kunita, M. Inami, T. Maeda, S. Tachi, “Real-Time Rendering System of Moving Objects,” Proc. IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes, pp. 81–88 (1999).
- [3] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video Based Modeling and Rendering of 3D Scenes,” IEEE Computer Graphics and Applications, Vol. 22, No. 2, pp. 66–73 (2002).

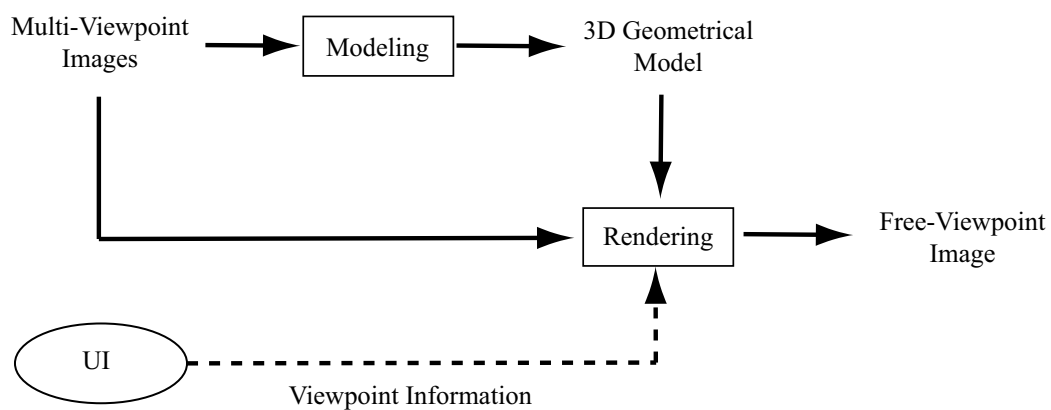


図 5.11: 従来の image-based modeling and rendering の処理フロー図 .

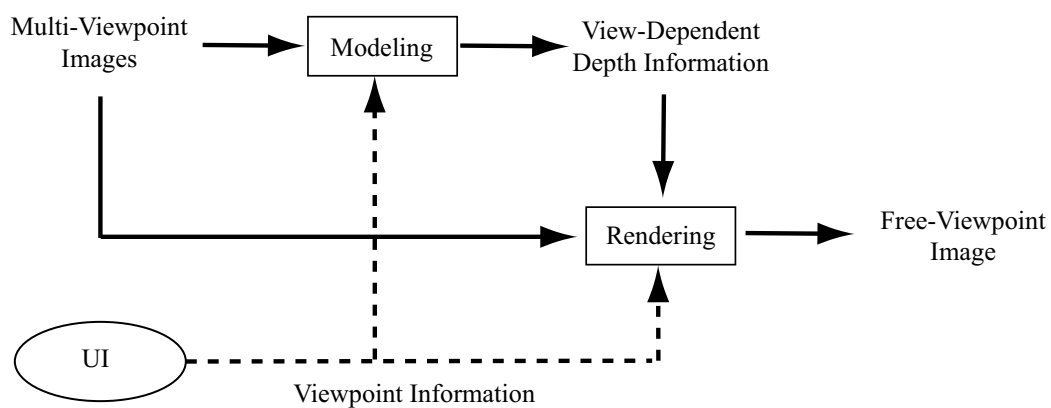


図 5.12: 視点依存型 image-based modeling and rendering の処理フロー図 .

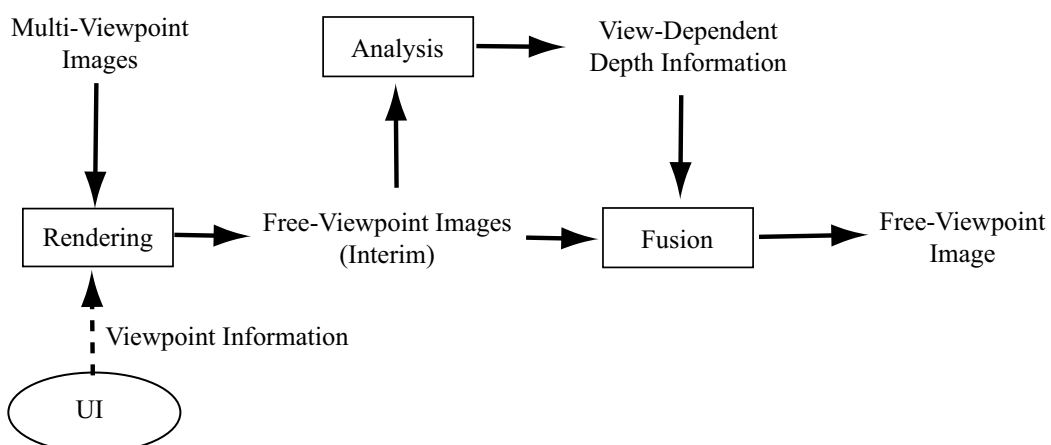


図 5.13: 本章の提案手法の処理フロー図 .

- [4] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera,” Proc. Eurographics Workshop on Rendering 2002, pp. 77–86 (2002)
- [5] R. Yang, G. Welch, G. Bishop, “Real-Time Consensus-Based Scene Reconstruction Using Commodity Graphics Hardware,” Proc. Pacific Conference on Computer Graphics and Applications, pp. 225–235 (2002).
- [6] P.N. Bangchang, T. Fujii, M. Tanimoto, “Experimental System of Free Viewpoint Television,” Proc. of SPIE 5006, Stereoscopic Displays and Virtual Reality Systems X, pp. 554–563 (2003).
- [7] G.G. Slabaugh, R.W. Schafer, M.C. Hans, “Image-Based Photo Hulls for Fast and Photo-Realistic New View Synthesis,” Real-Time Imaging 9, pp. 347–360 (2003).
- [8] C. Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” Proc. Eurographics Symposium on Rendering 2004, pp. 243–254 (2004)
- [9] T. Yamamoto, M. Kojima and T. Naemura, “LIFLET: Light Field Live with Thousands of Lenlets,” ACM SIGGRAPH 2004, Emerging Technologies, #0130 (2004)
- [10] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH 96, pp. 31–42 (1996).
- [11] S.J. Gortler, R. Grzeszczuk, R. Szeliski, M.F. Cohen, “The Lumigraph,” Proc. ACM SIGGRAPH 96, pp. 43–54 (1996).
- [12] J. -X. Chai, S. -C. Chan, H. -Y. Shum, and X. Tong, “Plenoptic Sampling,” Proc. ACM SIGGRAPH 2000, pp. 307–318 (2000)
- [13] J. Shade, S. Gortler, L. -W. He, and R. Szeliski: “Layered Depth Images,” Proc. ACM SIGGRAPH 98, pp. 231–242 (1998)
- [14] A. Isaksen, L. McMillan, and S.J. Gortler, “Dynamically Reparameterized Light Fields,” MIT-LCS-TR-778 (1999)
- [15] X. Tong, J.-X. Chai, and H. -Y. Shum, “Layered Lumigraph with LOD Control,” The Journal of Visualization and Computer Animation, Vol. 13, No. 4, pp. 249–261 (2002)
- [16] E. Adelson, J. Bergen, “The Plenoptic Function and the Elements of Early Vision,” in: M. Landy, J. Movshon (Eds.), Computer Models and Visual Processing, MIT Press, Cambridge (1991).
- [17] 藤井俊彰, “三次元統合画像符号化の基礎検討”, 博士論文, 東京大学大学院工学系研究科 電子工学専攻 (1994).

- [18] L. McMillan, G. Bishop, Plenoptic modeling: an image-based rendering system, Proc. ACM SIGGRAPH 95, pp. 39–46 (1995)
- [19] A. Isaksen, L. McMillan, and S.J. Gortler: “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH 2000, pp. 297–306 (2000)
- [20] R. Ng, “Fourier Slice Photography,” ACM Trans. on Graphics (Proc. ACM SIGGRAPH05), pp. 735–744 (2005).
- [21] S.K. Nayer and Y. Nakagawa, “Shape from Focus,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 16, No. 8, pp. 824–831 (1994).
- [22] S. Seitz and C. Dyer: “Photorealistic Scene Reconstruction by Voxel Coloring,” International Journal of Computer Vision, Vol. 25, No. 3, pp. 151–173 (1999)
- [23] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363 (1993)

第6章

合焦判定に基づく 全焦点自由視点画像合成 — 周波数領域における検討

6.1 まえがき

本論文では、カメラアレイによる多視点画像を入力とし、奥行きレイヤモデルを用いて自由視点画像合成を行う、という問題を考えている。第 5 章では、合成画像の各画素に適切な奥行きを割り当てるための新しい処理の枠組みとして、“合焦判定”と呼ばれる手法を導入した。第 5 章の提案は、空間領域における議論に基づくものであったが、本章では、光線情報の周波数解析に基づき、この“合焦”尺度に理論的な裏づけを与える。

本章の議論はシャノンのサンプリング理論に基づくものであるため、その適用範囲は光線情報が等間隔でサンプリングされるケースに限定される。これはすなわち、入力カメラに個体差がなく、各カメラが正確に平行かつ等間隔に配列されることを意味している。実世界においてこのような条件を実現するのは極めて困難であるが、本章の議論は、自由視点画像合成に対して、多次元信号処理に基づくユニークな裏づけを与える、一つの理論モデルとして価値があると考えられる。また、この“合焦”尺度を用いた“全焦点”画像合成の詳細な実験についても報告する。

6.2 背景

本節では、まず 6.2.1 で、本論文の第 5 章で提案した light field rendering における“合焦”判定法について述べる。次に 6.2.2 で、“合焦”判定法と他の奥行き推定の手法とを比較する。最後に 6.2.3 で、理論的背景として light field の周波数解析について述べる。

6.2.1 Light Field Rendering における“合焦判定”法

本章の議論の出発点は、イメージベースレンダリング [1] の中でも最も基本的な、light field rendering [2] と呼ばれる手法である。Light field rendering では、多数のカメラを密に配置して取得された多視点画像が入力として用いられる。それらの多視点画像はサンプリングされた光線の集合とみなされて、light field データとして光線空間 (Ray Space) でパラメータ化される。画像合成の際には、light field を補間・リサンプリングすることによって、所望の視点位置に到達する各光線を得る。この手法の大きな特徴は、被写体の明示的な構造を必要としないことにある。最も単純な場合、シーン全体が、focal plane と呼ばれる一枚の平面で近似される。

多視点画像が十分に高密度で取得されている場合、上記の近似によって、被写体全体を鮮明に捉えた自由視点画像を合成することができる。しかし、取得密度が不十分な場合、合成画像上では、focal plane の位置に応じて、一般のカメラの焦点と類似した現象が現われることが知られている [3, 4, 5]。すなわち、focal plane 付近の物体は鮮明に合成される（“焦点”が合う）が、focal plane から離れた物体には、ぼけや ghosting が発生する（“焦点”から外れる）¹。“焦点”が合う範囲、すなわち“被写界深度”は、入力画像の取得密度や、画像の解像度など複数のパラメータによって決まる [6, 7]。実際のシステムにおいては、この範囲が非常に狭い場合が多い、というのが問題となっている。

¹この現象を、一般の画像における焦点と区別するため、本章では、“焦点”のように二重引用符を用いて表記する。

Isaksen ら [3] は、上記の問題を解決するため、次のようなアプローチを提案した。

1. まず、通常の light field rendering 法で、focal plane の位置を変化させることで、“合焦” 距離が異なる多数枚の画像を合成する。
2. それらの画像の“焦点” が合っている領域を選択して、1 枚の画像に統合することで、“全焦点” 画像を得る。

(2) においては、画素ごとに、最適な focal plane の距離（以後、“合焦” 距離）を選ぶ処理が必要である。ある画素に対して、最適な“合焦” 距離を検出することは、その画素の奥行きを推定することと、おおよそ等価である。しかし、テクスチャの少ない領域については、実際の奥行きと、focal plane の奥行きがある程度違っている場合、合成画像上での見えに変化を生じないため、どの“合焦” 距離を最適とするかという点であいまいさが許容される。

画素ごとに“合焦” 距離を求める上記の処理において、Isaksen らは、あらかじめオフラインで独自の 3 次元モデル²を構築しておくことを前提とした手法を提案している。これに対して、本論文の第 5 章では、オフライン処理を行わず、(1) で合成した“多焦点” 画像を用いて、ランタイムで“合焦” 判定を行い、上記の処理を実現する手法を提案した。すなわち、オフライン処理を前提に、(1) の“多焦点” 画像を合成にのみ利用するのが Isaksen らの手法であり、オフライン処理を行わず、(1) の“多焦点” 画像を分析と合成に利用するのが提案手法である。提案手法のメリットは、オフライン処理を必要としないため、動的なシーンを扱うのにより適している点である。今後は、ライブの自由視点画像合成システム [8, 9] への適用も期待される。

6.2.2 他の奥行き推定法との比較

多数枚の画像を利用して、対象空間の奥行きや被写体の形状を推定する手法は、大きく二つに分けられる。一方は、多視点画像を用いるアプローチで、これには stereo 法 [10] と silhouette 法 [11] が含まれる。他方は、多焦点画像を用いるアプローチで、depth-from-focus 法と呼ばれる [12]。提案手法では、多視点画像から合成された“多焦点” 画像を用いて、“合焦” 判定を行い、空間の奥行きを推定する。したがって、本手法は、上記の 2 つのアプローチの両方の要素を備えているとみなすことができる。

提案手法の長所として、任意の視点位置における距離画像を直接得ることができる、という点がある。一般的な depth-from-stereo 法や depth-from-focus 法では、距離画像の視点は、実際に画像が撮影されたカメラ位置に限定される。新しい視点位置から見た距離画像を得るためには、3 次元の座標変換に加えて、オクルージョン領域を考慮した複雑な処理が必要である [13]。しかし、提案手法では、ユーザが指定した任意の視点位置で合成された画像群が用いられるため、その視点位置に対応した距離画像を、直接得ることができる。これは、自由視点画像合成の必要を満たす性質である。

提案手法は、“多焦点” 画像に対する“合焦” 判定を行うので、処理の手順という点では depth-from-focus 法に類似する。しかし、判定に用いる“合焦” の尺度において相違がある。depth-from-

²多層レイヤモデルにおいて、各層上の各点に被写体が存在する確からしさが与えられたもの。

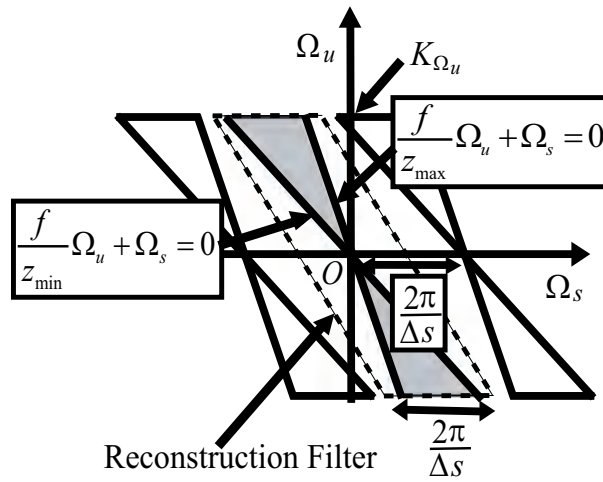


図 6.1: 多視点画像信号の周波数解析.

focus 法では，高周波成分が合焦の尺度として用いられる [12]．これは，一般の画像においては，焦点が合っている領域ほど，より多くの高周波成分を持つためである．しかし，提案手法では，light field rendering による合成画像が処理の対象となり，その“焦点”の性質は，一般の画像とは異なっている．合成画像上の“焦点”が合っていない領域には，ぼけだけではなく，高周波を含む ghosting 雑音が発生する．したがって，高周波成分を用いた一般的な合焦尺度は，light field rendering による合成画像に対しては十分に機能しない．そこで，本論文では，light field rendering における“焦点”を安定に判定するため，数種類の再構成フィルタを組み合わせる新しい“合焦”尺度を提案する．

6.2.3 Light field の周波数解析と“焦点”

入力として用いられる多視点画像は，2次元格子状に平行に並べられた多数のカメラによって取得されるものとする．カメラの位置を (s, t) ，各カメラで撮影される画像上の画素の位置を (u, v) と表すことにより，4次元の light field (s, t, u, v) が定義される．取得された各光線は，4つのパラメータの組 (s, t, u, v) によって，一意に識別される．以下，簡単のため4次元空間 (s, t, u, v) のうち，部分空間 (s, u) について議論する． (s, u) 空間のフーリエ変換領域を， (Ω_s, Ω_u) と表記する．

Chai ら [14] は，light field の信号特性を扱いやすいものとするため，対象とするシーンについて，1) 被写体の反射特性はランバートとして扱える，2) オクルージョンの影響が無視できる，という2点を仮定した．この条件下においては，信号のスペクトラムは，図 6.1 の影をつけた領域に制限される（導出の過程などの詳細は文献 [14] を参照されたい）．ここでは，対象とする空間の奥行きが $z_{\min} \leq Z \leq z_{\max}$ で定められるとする． f はカメラの焦点距離を表す． K_{Ω_u} は， u 軸方向の周波数の最大値を表し，シーンの模様の複雑さやカメラおよび合成画像の解像度によって決まる．

Light field は一定のカメラ間隔 Δs で離散的にサンプリングされるため，周波数領域では Ω_s 軸

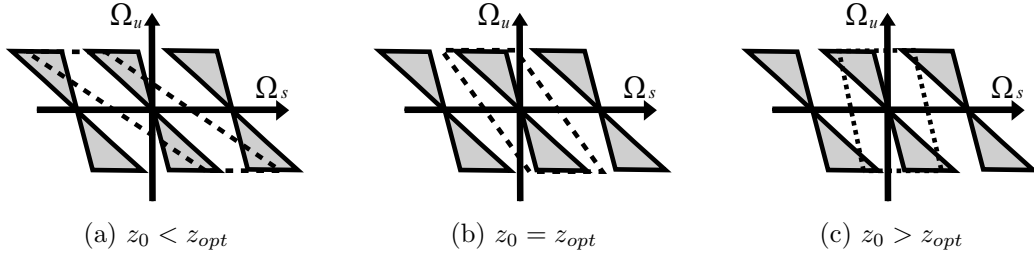


図 6.2: 仮定される奥行きと再構成フィルタ.

の方向に $2\pi/\Delta s$ の周期で原信号の複製成分が繰り返される．図 6.1 では，左右の隣接する複製成分の輪郭を図示した．複製成分が元の信号と重なり合わないためには，下記の条件が満たされる必要がある．

$$\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \leq \frac{2\pi}{K_{\Omega_u} f \Delta s} \quad (6.1)$$

画像合成の際には，離散的にサンプリングされた光線情報を補間する必要がある．これは，ローパスフィルタによって，連続信号を再構成することと等価である．Chai ら [14] は，図 6.1 に示すような，幅 $2\pi/\Delta s$ の平行四辺形状のローパスフィルタを提案した．このフィルタによって，スペクトラムは，平行四辺形の内部の領域に帯域制限される．このとき平行四辺形の傾きは，画像合成の際に focal plane を仮定する奥行き z_0 に対応する（傾き： $-z_0/f$ ）． z_0 の最適値 z_{opt} は下記のように定められる．

$$\frac{1}{z_{opt}} = \frac{1}{2} \left(\frac{1}{z_{\min}} + \frac{1}{z_{\max}} \right) \quad (6.2)$$

以上の議論を元に，light field rendering における“焦点”の現象を説明することができる [4, 5]．多くの場合，カメラ間隔 Δs が大きいため，対象とするシーン全体に対して (6.1) 式を満たすのは困難である．しかし，シーン中の比較的狭い範囲に注目すると，奥行きの変化の幅（ $(1/z_{\min} - 1/z_{\max})$ に対応）が十分に小さく，(6.1) 式が満たされる領域があると考えられる．このような領域を対象として，ある奥行きを仮定したときの，再構成フィルタの様子を図 6.2 に示す． $z_0 = z_{opt}$ のとき（図 6.2(b)）には，光線信号は，過不足なく適切に再構成される（“焦点”が合う）．しかし， $z_0 < z_{opt}$ や $z_0 > z_{opt}$ のとき（図 6.2(a), (c)）には，再構成において，原信号の高周波成分が漏れたり，近傍の折り返し成分が含まれたりする場合があることが分かる（“焦点”から外れる）．これらは，合成画像において，ぼけ（高周波成分の漏れに対応）や ghosting（折り返し成分に対応）が発生する原因である．

Zhang ら [15] は，Chai ら [14] の仮定を緩め，被写体空間にオクルージョンや非ランバート反射が存在する場合に理論を拡張した．結論として，これらの成分が帯域制限されている場合は，十分な干渉対策用帯域（ガードバンド）を設けることにより，同様の再構成フィルタが適用可能であることが示された．ガードバンドを設けるためには，(6.1) 式で等号が成立する場合よりも，カメラの間隔 Δs を小さくするか，空間周波数 K_{Ω_u} を小さくする（解像度を下げる）必要がある．

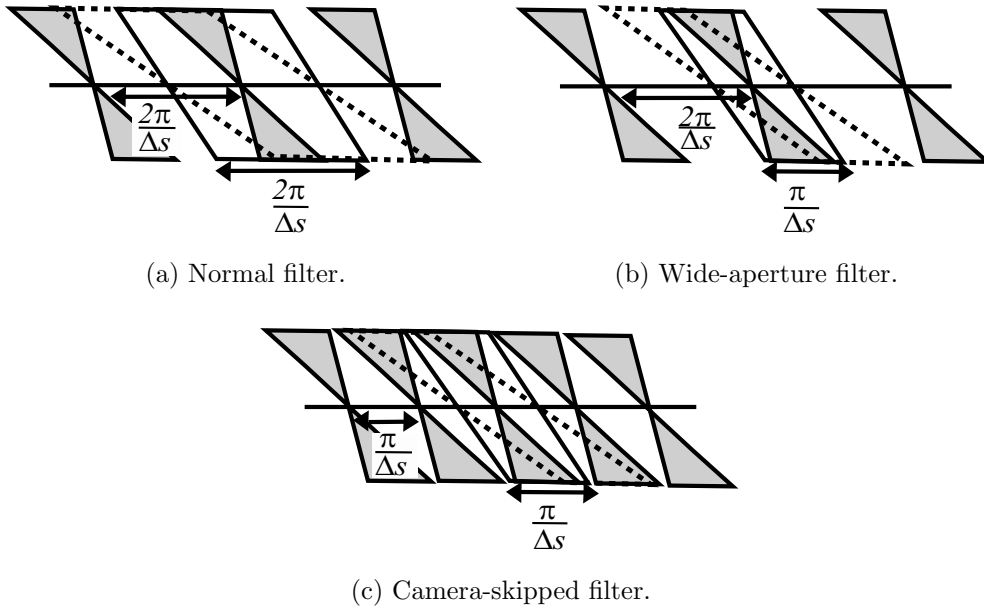


図 6.3: 周波数領域における各フィルタの形状 .

以後、本論文では、Chai ら [14] の仮定が成り立つ範囲で議論を進めることにするが、Zhang ら [15] の議論に結びつけることが可能である .

6.3 提案手法

6.3.1 Light Field Rendering における合焦判定の原理

Light field rendering による合成画像における“焦点ぼけ”は、一般のカメラ画像のものとは性質が異なり、折り返し雑音による高周波成分を含んでいる . したがって、depth-from-focus 法で用いられるような、高周波成分の大きさに基づいた“合焦”尺度を適用するのは困難である . そこで本章では、Light field rendering による合成画像に特化した、合焦判定の尺度を提案する . 本手法では、高周波成分ではなく、何種類かの再構成フィルタの差異を“合焦”判定に利用する .

まず、再構成フィルタの種類による合成結果の違いについて、周波数領域で説明を行う . ここでは、シーン全体のうち、比較的狭い領域について考える . その領域については、奥行きの変化の幅 ($(1/z_{\min} - 1/z_{\max})$ に対応) が十分に小さく、与えられたカメラ間隔 Δs に対して、

$$\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \leq \frac{1}{2} \cdot \frac{2\pi}{K_{\Omega_u} f \Delta s} \quad (6.3)$$

が満たされると仮定する . (6.1) 式よりもさらに強い条件となっているのは、Chai ら [14] の提案したフィルタに加えて、他の種類のフィルタも併用するためである .

本手法では、図 6.3 に示した三種類の再構成フィルタを用いる .

normal filter : 最も基本的なフィルタで、幅は $2\pi/\Delta s$ である . $z_0 = z_{opt}$ のとき、Chai ら [14] の示した最適フィルタになる .

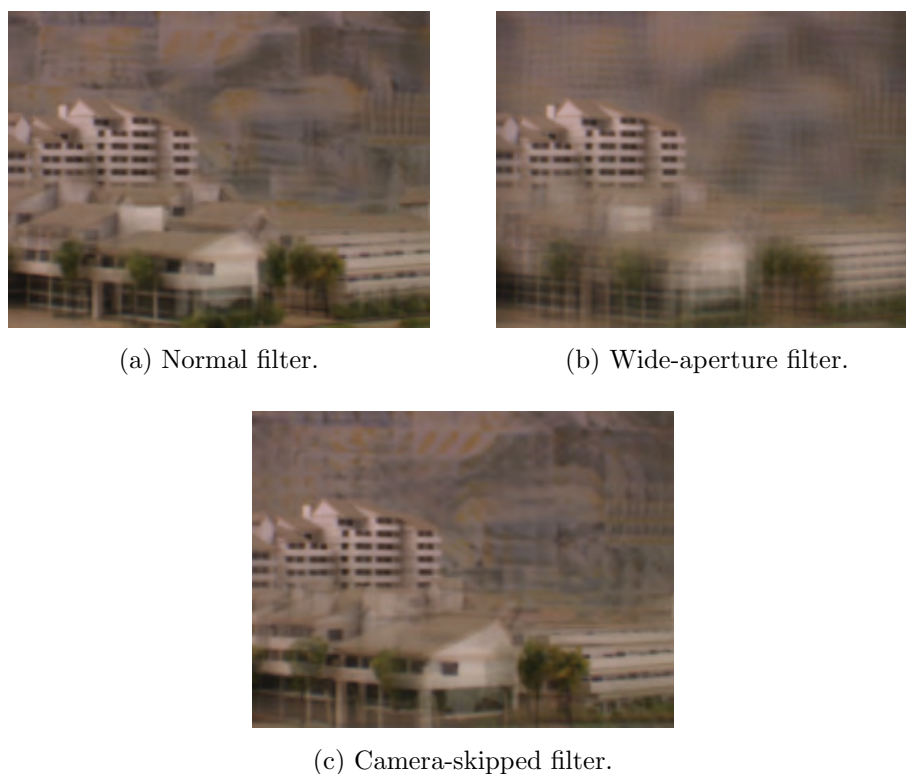


図 6.4: 異なる再構成フィルタによる合成画像 .

wide aperture filter : このフィルタの幅は, $2\pi/\Delta s$ よりも小さい . 開口 (aperture) が広いカメラで撮影したような合成結果が得られることから, このように名づけられている [3, 4, 5] . 本論文では, 図 6.3(b) に示したようにフィルタの幅を $\pi/\Delta s$ とする .

camera-skipped filter : 入力画像をひとつおきに間引いた信号に対して, normal filter を施す . 信号の繰返し周期, フィルタの幅が共に $\pi/\Delta s$ となる .

ここに挙げた 3 種類のフィルタは, いずれも normal filter の自然な拡張であり, 空間波形が単純なため, 簡便かつ高速にフィルタ処理を実装できる (詳細は付録 B を参照されたい) . 本稿では議論しないが, より複雑なフィルタについても研究されている [15] .

図 6.3 において, 実線で示されているのは, $z_0 = z_{opt}$ (“焦点” が合う) の場合である . このとき各再構成フィルタを通過するスペクトラムは, 全く同一となる . したがって理論的には, すべてのフィルタが全く同一の合成結果を与えることになる . 一方, 図中に破線で示されているのは, $z_0 < z_{opt}$ の場合である . この例に示されるように, $z_0 \neq z_{opt}$ (“焦点” から外れる) の場合, 通過するスペクトラムはフィルタごとに異なる . したがって, 合成結果もフィルタごとに異なったものとなると考えられる .

図 6.4 に, 各フィルタを用いて合成した自由視点画像の例を示す . 入力として, 「筑波大学多視点画像データベース」の画像群を使用した . 上記の議論は, 図 6.4 の合成結果と一致するもので

ある．これらの画像は，同一の視点位置，同一の“合焦”距離において合成されている．“焦点”が合っているのは，いずれも左側の奥に配置されている高層ビルの部分である．いずれのフィルタによっても，“焦点”が合っている領域はほとんど同じように合成されるが，“焦点”から外れた領域におけるぼけや ghosting の発生の仕方は，フィルタごとに差異があることが分かる．

そこで筆者らは，同一の“合焦”距離で異なる再構成フィルタによって合成された画像間の差分を，light field rendering における“合焦”尺度として用いることを提案する．これは，ある画素が“合焦”となるとき，その画素における差分の絶対値は，比較的小さい（理論的には 0）という性質に基づく．シーンのテクスチャの変化が少ない領域などにおいては，“焦点”が合っていない画素が，0 に近い差分値を取る場合もあると予想される．しかし，このような領域は，その奥行きが正しく推定されなくても，最終的な合成画像の品質に影響を与えない．提案手法の目的は，高精度な奥行き推定ではなく，高品質な合成画像を得ることである．

6.3.2 “合焦”判定による奥行き推定と“全焦点”画像合成

あらかじめ， N 個の“合焦”距離の候補， z_n ($n = 0, 1, 2, \dots, N-1$) を定める．本章では，視差空間を等分するように，下記の式に従うことにする．

$$\frac{1}{z_n} = \frac{n}{N-1} \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) + \frac{1}{Z_{\max}} \quad (6.4)$$

ただし， Z_{\max} ， Z_{\min} は，シーン全体の奥行きの最大値/最小値を表す．また，三種類の再構成フィルタ，normal filter, wide-aperture filter, camera-skipped filter をそれぞれ h_1 ， h_2 ， h_3 と表記する．

まず，ユーザに与えられた視点位置に対して，“多焦点”画像を合成する．フィルタ h_k によって，“合焦”距離 z_n で合成された画像を $I_k^{(n)}(x, y)$ と表す． (x, y) は画素位置を表す．本論文では三種類の再構成フィルタを用いるため，同一の視点位置における自由視点画像を，合計で $3N$ 枚合成する．

提案手法では，再構成フィルタの差異を“合焦”判定に利用する．そこで，まず同一の z_n において異なるフィルタによって合成された画像間の差分を取る．

$$Sub^{(n)}(x, y) = \sum_{k < l} a_{k,l} \left| I_k^{(n)}(x, y) - I_l^{(n)}(x, y) \right| \quad (6.5)$$

ここで， $a_{k,l}$ は重み付け係数である．フィルタの種類は最低 2 つで十分であるが，3 種類以上を用いることによって，“合焦”判定をより安定化できる．さらに， $Sub^{(n)}(x, y)$ を， $(2M+1)^2$ のプロック領域内で平均値を取ることで安定化を図る．

$$F^{(n)}(x, y) = \frac{1}{(2M+1)^2} \times \sum_{-M < i, j < M} Sub^{(n)}(x+i, y+j) \quad (6.6)$$

こうして得られた $F^{(n)}(x, y)$ を，点 (x, y) における“合焦”尺度として定義する．

点 (x, y) に対する，最適な“合焦”距離のインデックス $n(x, y)$ は，“合焦”尺度 $F^{(n)}(x, y)$ を最小化する n として，次のように求められる．

$$n(x, y) = \arg \min_{0 \leq n \leq N-1} F^{(n)}(x, y) \quad (6.7)$$

すべての点 (x, y) について $n(x, y)$ を求めることにより，所望の視点位置から見た距離画像が得られる．最後に，点 (x, y) ごとに，最適な“合焦”距離 $n(x, y)$ における合成画像 $I_1^{(n(x, y))}$ から画素値を読み出すことにより，“全焦点”画像 $I(x, y)$ が得られる³．

$$I(x, y) = I_1^{(n(x, y))}(x, y) \quad (6.8)$$

6.4 実験

実験には，Linux ベースの汎用の PC (CPU: Pentium 4 2.0GHz, Memory: 2GB, GPU: Geforce4 Ti 4200) を利用した．すべてのプログラムは C と OpenGL で記述し，gcc でコンパイルした．再構成フィルタは，本来は理想ローパスフィルタであるが，実装上の便宜および十分な合成結果が得られることから，線形補間で近似することにした．数学的詳細については，付録 B を参照されたい．線形補間を用いた場合，normal filter および camera-skipped filter は，それぞれ，第 5 章で導入した normal mode および reference mode と等価になる．ただし，第 5 章の議論は任意のカメラアレイに適用可能であったのに対し，本章の議論は，カメラ間に個体差がなく正方格子状にアレイが構成される，理想的なケースに限定される．

6.4.1 実写画像を用いた実験

まず，「筑波大学多視点画像データベース」に含まれる，81 視点（縦 9 × 横 9）の実写画像を用いて，画像合成の実験を行った（図 6.5）．入力多視点画像の視点間隔は 8 mm，画素数 640 × 480，水平視野角 27° である．GPU 上で処理を行うため，これらの画像は 256 × 192 画素に縮小して利用した．被写体となっているのは街の模型である．カメラシステムから，最も手前の建物までの距離は約 300mm，最も遠い建物までの距離は約 600 mm，背景面までの距離は 1840 mm である．合成画像の解像度は 512 × 512，画角は 27° とした⁴．(6.4) 式においては，カメラの画角などのパラメータの誤差を考慮に入れ， $Z_{\max} = 1840$ mm， $Z_{\min} = 250$ mm とした．また，合焦距離の候補数 $N = 14$ ， $a_{1,2}, a_{2,3} = 1, a_{1,3} = 0$ ((6.5) 式)， $M = 7$ ((6.6) 式) とした．

図 6.6 – 6.8 に，3 種類の再構成フィルタによる，“合焦”距離が異なる合成画像の例を示す．これらの画像は，中央の入力画像が取得された位置から，手前に 200 mm 引いた視点位置で合成されたものである．“合焦”距離は， n が 0 から 13 へと増加するのに伴って，奥から手前へと移動している．提案手法では，これらの 3 つの“多焦点”画像系列を互いに比較することによって“合焦”判定を行い，“焦点”が合っている部分を統合することで，“全焦点”画像を合成する．

“合焦”判定の様子を，図 6.9(a) に示した 2 点 (A, B) を用いて説明する．図 6.9(b) は，奥行きインデックス n に対して，(6.6) 式によって定義される“合焦”尺度 $F^{<n>}$ の値をプロットしたものである．点 A は $n = 4$ ($Z = 622$ mm)，点 B は $n = 9$ ($Z = 341$ mm) で，“合焦”尺度が最小となり，最も“焦点”が合う，と判定される．これらの Z の値は，いずれも，実際のシーンとおおよそ一致するものである．図 6.6 – 6.8 から，それぞれの点が，最適と判定された“合

³ $I_2^{(n(x, y))}$, $I_3^{(n(x, y))}$ を用いてもよい．

⁴ 結果を見やすくするため，本節では，合成画像の一部を拡大して掲載する．



図 6.5: 入力多視点画像（「筑波大学多視点画像データベースより」）。

焦”距離において，実際に鮮明に合成されていることがわかる．ほとんどの画素については，上記の例と同様， $F^{<n>}$ の最小値を安定に求めることができる．

図 6.10 に，すべての画素について同様の“合焦”判定処理を行って得られた距離画像を示す．手前に位置する画素ほど，高い輝度が割り当てられている．被写体の奥行き構造が，おおよそ正しく推定されているものの，特に，手前の物体と奥の物体の境界付近には，奥行きが正しく判定されていない部分もあることがわかる．これは，奥行きの変化が不連続になる領域では，(6.3) 式が満たされないためと考えられる．これらの誤差は，何らかの平滑化処理を行うことで低減できると考えられるが，本章では処理速度を優先するためこの処理を割愛している．また，奥行き推定そのものが目的ではないため，距離画像の精度は議論しない．

図 6.11 に，最終的な“全焦点”画像を示す．図 6.12 に示したように，他の任意の視点位置においても，上記と同様の処理を行うことによって，“全焦点”画像を合成することができる．従来の light field rendering（例えば図 6.6）では，特定の奥行きだけに“焦点”が合った画像しか合成できなかったが，提案手法では，シーン全体に“焦点”が合った，視覚的に良好な画像が合成されていることが分かる．

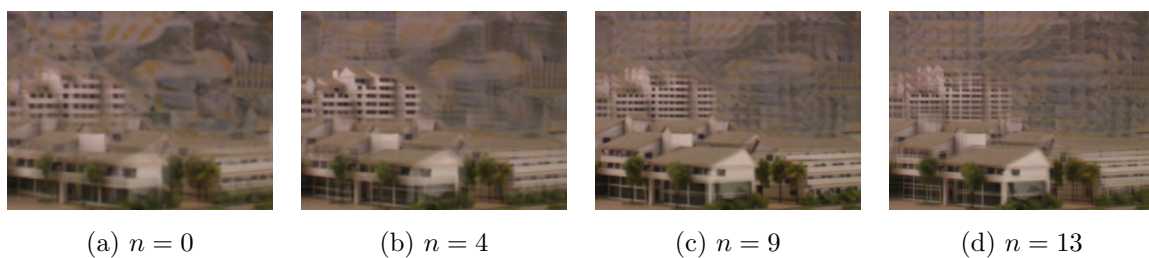


図 6.6: “合焦” 距離が異なる合成画像 (normal filter による) .

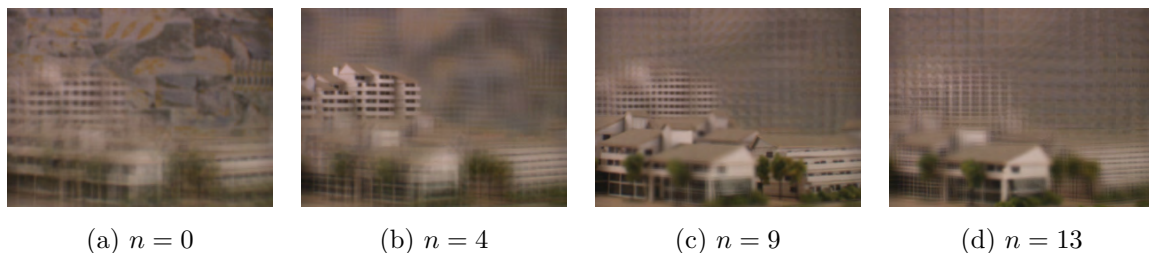


図 6.7: “合焦” 距離が異なる合成画像 (wide-aperture filter による) .

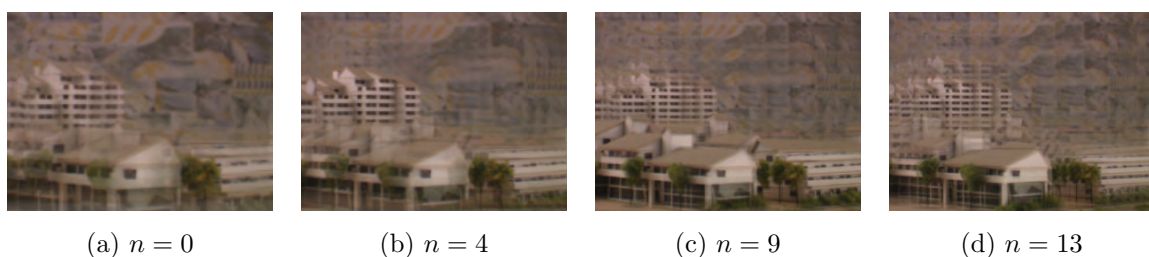


図 6.8: “合焦” 距離が異なる合成画像 (camera-skipped filter による) .

6.4.2 定量的な評価

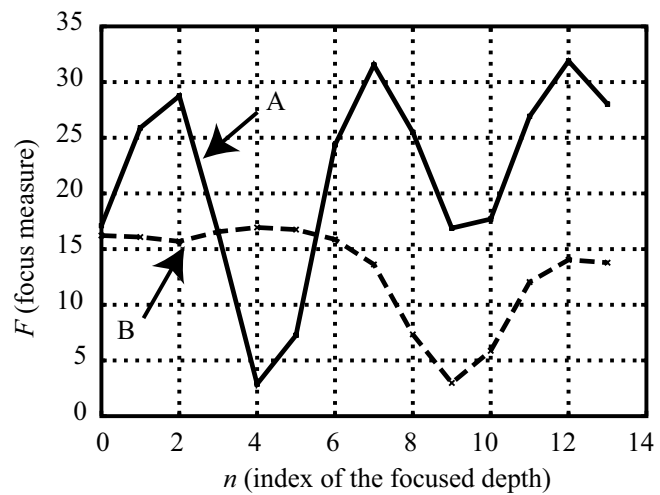
提案手法による合成品質を定量的に評価するため、CG シーンを利用した画像合成の実験を行った。本実験では、まず CG シーンを 9×9 の 81 視点においてレンダリングし、画面をキャプチャすることで、入力として用いる多視点画像群を取得する。次に、提案手法によって、新たな視点で距離画像および“全焦点”画像を合成する。最後に、最終的な合成画像と、元の CG シーンを比較して PSNR を算出する。実験のパラメータの詳細は、表 1 に示した通りである。 Z 軸は、入力画像群を取得した平面に直交するように定義されている。

実験に使用したシーンは、6 枚の平面で構成され、中央が奥に“凹んだ”構造になっている。表面にはテクスチャがマッピングされている。シーンの構造および外観を示すため、図 6.13(a) に推定された距離画像、(b) に合成された“全焦点”画像の例を挙げる。本実験では、奥行きの変化が連続的になるようなシーンを想定した。また、見かけ上の PSNR が高くなるのを防ぐため、合成画像の画面内に単色の背景領域が入らないようにも配慮した。

画像合成の際の focal plane の配置は (6.4) 式に従うが、 $N = 1$ の場合には、(6.2) 式によって



(a) 注目する2点(画像は図6.6(b)と同一)



(b) “合焦”尺度の値の変化

図 6.9: 合焦判定による奥行き推定 .

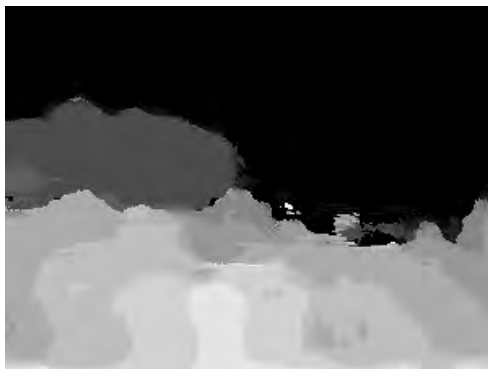


図 6.10: 推定された距離画像 .



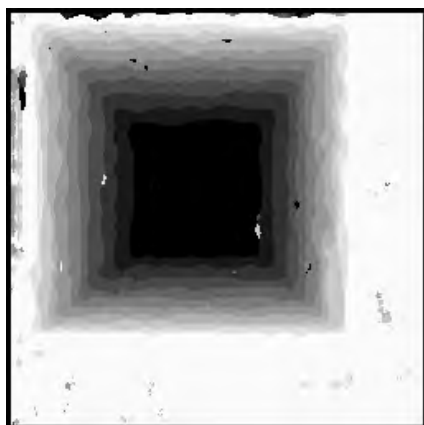
図 6.11: 提案手法による“全焦点”画像 .



図 6.12: 他の視点に対する“全焦点”画像 .

表 6.1: CG を用いた実験の諸条件

CG scene	$10 \leq Z \leq 20$
input images	$Z = 0$ interval: 0.9, number: 9×9 pixels: 256×256 , FOV: 27°
synthetic images	$Z = -20$ pixels: 256×256 , FOV: 12°



(a) depth map



(b) synthetic image

図 6.13: CG シーンを用いた合成 ($N = 11$) .

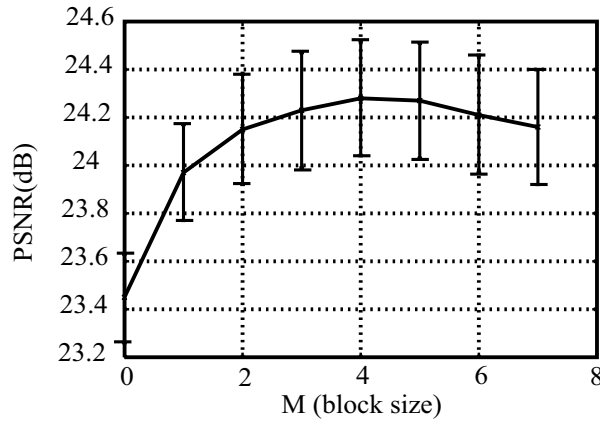
定められる最適値とする．提案手法において変数となるのは，

1. “合焦” 距離の候補数 N .
2. 再構成フィルタ間の差分に対する重み付け係数 ((6.5) 式の $a_{k,l}$) .
3. 平均値処理のブロックサイズ ((6.6) 式の M) .

の 3 つである．これらの変数の値を変化させ，最終的な合成画像 (“全焦点” 画像) の PSNR を比較した．

まず，(6.6) 式における平均値処理のブロックサイズ M を変化させ，合成される “全焦点” 画像の PSNR を比較した．この実験では， $N = 11$ ， $a_{1,2} = 1$ ， $a_{1,3} = 1$ ， $a_{2,3} = 0$ とした．それぞれの M について，入力画像の視点とは異なる 81 視点で画像を合成し，それらの画像の PSNR の平均値，および標準偏差を図 6.14 に示した⁵．最適な M の値は，合成画像の解像度やシーンの複雑さに依存すると考えられるが，今回利用したシーンに関しては， $M = 4$ で PSNR が最大となった．しかし， M が大きくなると計算量も増加するため，妥当な合成品質が得られる範囲で，なるべく小さな値を設定することが望ましい．そこで，以後の実験では， M の値は 3 に固定することにした．

⁵合成画像の PSNR は視点位置に依存するため，統計的な値で評価することが必要である．81 の視点は $Z = -20$ において，間隔 0.1， 9×9 の 2 次元格子上に配置されている．

図 6.14: 平均値処理のブロックサイズ M と PSNR の関係表 6.2: 係数 $a_{k,l}$ の組み合わせ

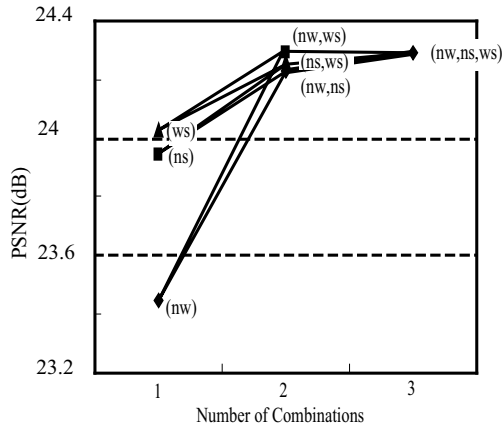
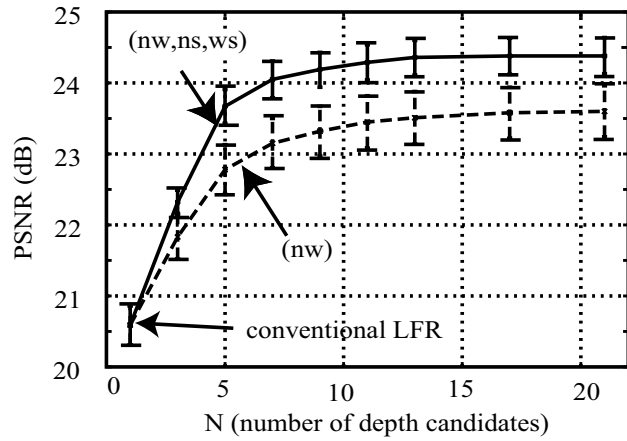
	$a_{1,2}$	$a_{1,3}$	$a_{2,3}$	差分の組み合わせ
(nw,ns,ws)	1	1	1	3
(nw,ns)	1	1	0	2
(nw,ws)	1	0	1	
(ns,ws)	0	1	1	
(nw)	1	0	0	1
(ns)	0	1	0	
(ws)	0	0	1	

次に, (6.5) 式の係数 $a_{k,l}$ の値を変化させて, 合成画像の PSNR を計測した. $N = 11$ とした. 本実験では, $a_{k,l}$ の値を下記の 2 値のいずれかとする.

$$a_{k,l} = \begin{cases} 1 & : h_k \text{ と } h_l \text{ の差分を考慮する} \\ 0 & : h_k \text{ と } h_l \text{ の差分を考慮しない} \end{cases} \quad (6.9)$$

$a_{k,l}$ の組み合わせを表 2 に示す. ラベルの n,w,s はそれぞれ, normal filter, wide-aperture filter, camera-skipped filter を意味する. (nw,ns,ws) では, “合焦” 判定の際に, 3 通りすべての差分の組み合わせが考慮されるのに対して, (nw,ns), (nw,ws), (ns,ws) では 2 通り, (nw), (ns), (ws) では 1 通りのみが考慮される. 図 6.15 に, それぞれの組み合わせによる合成画像の PSNR (上記と同じ 81 視点に対する平均値) を示す. 全体的な傾向として, “合焦” 判定の際に考慮する差分の組み合わせを 1 から 2 以上に増やすことにより, 合成画像の PSNR を高められることがわかる.

最後に, “合焦” 距離の候補数 N を変化させて, 合成画像の PSNR を測定した. 表 2 の (nw,ns,ws) および (nw) の 2 つの場合についての結果を図 6.16 に示した. 上記と同様に, それぞれの N について, 81 視点で合成を行い, 平均値と標準偏差を示している. $N = 1$ の場合の合成結果は, 従来の light field rendering で focal plane を最適な位置に配置した場合と等価である. グラフから, (nw,ns,ws), (nw) いずれの場合も, N の増加に伴って合成画像の PSNR が単調に増加し, 徐々に

図 6.15: 係数 $a_{i,j}$ の組合わせと合成品質図 6.16: 合焦距離の候補数 N と PSNR の関係

飽和する傾向があることがわかる．(nw,ns,ws) の場合，従来の light field rendering と比較して PSNR が平均値が 3.8 dB 向上した．これらは提案手法の有効性を支持する結果である．

以上の定量的な実験の結果をまとめると，“合焦”距離の候補数 N が大きいほど，また再構成フィルタ間の差分の組み合わせが多いほど，最終的な合成画像の品質が高くなること分かる．しかし， N の増加およびフィルタ間の差分の組み合わせ数の増加は，いずれも計算のコストを増大させる．特に N は支配的であり，計算時間はおおよそ N に比例する．例えば，(nw,ns,ws) の場合，1 枚の“全焦点”画像の合成（“多焦点”画像の合成も含む）に要する時間は， $N = 5$ では 106 msec， $N = 11$ では，235 msec であった．したがって，実用に際しては，合成品質と処理時間のトレードオフを考慮して，適切な N の値を選択しなければならない．

6.5 むすび

本章では，第 5 章で提案した“合焦”判定に基づく“全焦点”自由視点画像合成手法に対して，光線情報の周波数解析に基づく新たな理論的考察を導入した．まず，周波数解析に基づき，light field rendering における“焦点”の現象を説明し，“合焦”領域を判定する尺度の議論と結びつけた．また，その尺度を用いて，“合焦”距離が異なる合成画像群の“焦点”が合った領域を統合し，“全焦点”画像を合成する手法を示した．実験によって，視覚的，また定量的に，最終的な合成画像の品質が向上することが確認され，提案手法の有効性が示された．今後は，インタラクティブな画像合成を可能にするため，画質と処理時間のトレードオフを検討するとともに，高速実装技術についても研究を行う．また，本章の議論を拡張し，コンピュータビジョン分野の形状推定手法と再構成フィルタの関係について理論的考察を進める．

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. Circuits and Syst. for Video Tech.*, Vol. 13, No. 11, pp. 1020–1037 (2003).
- [2] M. Levoy and P. Hanrahan, “Light Field Rendering,” *Proc. ACM SIGGRAPH 96*, pp. 31–42 (1996).
- [3] A. Isaksen, M. Leonard, and S.J. Gortler, “Dynamically Reparameterized Light Fields”, Technical Report MIT-LCS-TR-778 (1999).
- [4] A. Isaksen, L. McMillan, and S. J. Gortler, “Dynamically Reparameterized Light Fields”, *Proc. ACM SIGGRAPH 00*, pp. 297–306 (2000).
- [5] J. Stewart, J. Yu, S.J. Gortler, and L. McMillan, “A New Reconstruction Filter for Under-sampled Light Fields,” *Proc. Eurographics Symposium on Rendering*, pp. 150–156 (2003).
- [6] 國田 豊, 稲見昌彦, 前田太郎, 館 日章, “多眼カメラを用いた任意視点人物像の実時間生成システム”, *信学論* Vol. J84-D-II, No. 1, pp. 129–138 (2001).
- [7] 高橋桂太, 苗村 健, 原島 博, “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の理論”, *映情学誌*, Vol. 57, No. 10, pp. 1292–1299 (2003).
- [8] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video-Based Modeling and Rendering of 3D Scenes,” *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, pp. 66–73 (2002).
- [9] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera,” *Proc. 13th Eurographics Workshop on Rendering*, pp. 77–86 (2002).
- [10] U.R. Dhond and J.K. Aggarwal, “Structure from stereo – a review,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 6, pp. 1489–1510 (1989).
- [11] A. Laurentini, “The Visual Hull Concept for Silhouette-Based Image Understanding,” *IEEE Trans. Pattern Anal. & Mach. Intel.*, vol. 16, no. 2, pp. 150–162 (1994).
- [12] S.K. Nayer and Y. Nakagawa, “Shape from Focus,” *IEEE Trans. Pattern Anal. & Mach. Intel.*, Vol. 16, No. 8, pp. 824–831 (1994).
- [13] 昼間香織, 沖村隆幸, 上平員丈, “複数の奥行マップを用いた仮想視点画像の生成”, *信学論*, Vol. J84-D-II, No. 5, pp. 805–811 (2001).
- [14] J.-X. Chai, X. Tong, S.-C. Chany, and H.-Y. Shum, “Plenoptic Sampling,” *Proc. ACM SIGGRAPH*, pp. 307 – 318 (2000).

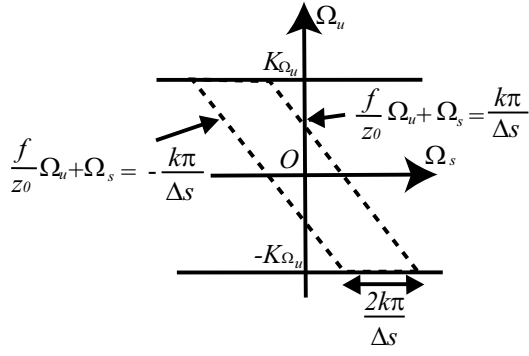


図 6.17: 周波数領域におけるフィルタ .

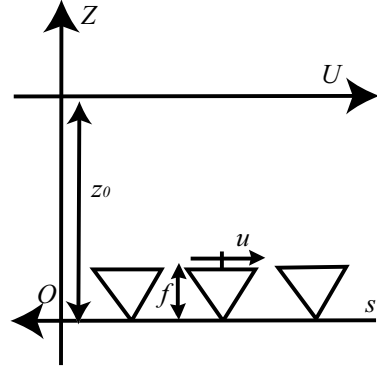


図 6.18: U と s-u の関係 .

- [15] C. Zhang and T. Chen, “Spectral Analysis for Sampling Image-Based Rendering Data”, IEEE Trans. Circuits and Systems for Video Technology, Vol. 13, No. 11, pp. 1038–1050 (2003).

6.6 付録

6.6.1 再構成フィルタの空間波形の導出

再構成フィルタの通過帯域は，図 6.17 に示した平行四辺形の内部の領域，すなわち，

$$\begin{aligned} -\frac{k\pi}{\Delta s} < \frac{f}{z_0}\Omega_u + \Omega_s < \frac{k\pi}{\Delta s} \\ -K_{\Omega_u} < \Omega_u < K_{\Omega_u} \end{aligned} \quad (6.10)$$

が満たされる領域である．図 6.3 に示したように，normal filter では $k = 1$ ，wide-aperture filter および camera-skipped filter では $k = 1/2$ となる．ゲイン G の理想的なローパスフィルタを仮定すると，周波数特性は次のように表される．

$$H(\Omega_s, \Omega_u) = \begin{cases} G & \text{(6.10) 式} \\ 0 & \text{otherwise} \end{cases} \quad (6.11)$$

再構成フィルタの空間波形 $h(s, u)$ は， $H(\Omega_s, \Omega_u)$ の逆フーリエ変換として，次のように求められる．

$$\begin{aligned} h(s, u) &= \mathcal{F}^{-1}[H(\Omega_s, \Omega_u)] \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H e^{j(\Omega_s s + \Omega_u u)} d\Omega_s d\Omega_u \\ &= A \cdot Sa\left(\frac{k\pi}{\Delta s} s\right) Sa\left(K_{\Omega_u} \left(u - \frac{f}{z_0} s\right)\right) \end{aligned} \quad (6.12)$$

ただしここで， A は，

$$A = \frac{1}{\pi^2} \cdot \frac{k\pi}{\Delta s} \cdot K_{\Omega_u} \cdot G \quad (6.13)$$

なる定数, $Sa(x)$ は標準化関数

$$Sa(x) = \frac{\sin(x)}{x} \quad (6.14)$$

である．以後，簡単のため定数 A は除外して考える．

6.6.2 近似的な実装

画像合成を実装するには，奥行き $Z = z_0$ の位置に置かれた focal plane によって，シーン全体を近似する．図 6.18 のように，focal plane 上に U 軸を定義する．画像合成の際には， (s, U) 空間で，必要な光線の補間合成が行われる，と考えることができる．ここで， U 軸の原点を Z 軸との交点， u 軸の原点を各カメラの撮像面の中心とすると， U と s, u の間には，次の関係式が成り立つ．

$$U = -s + \frac{z_0}{f}u \quad (6.15)$$

このとき，(6.12) 式によって得られたフィルタの波形を U を用いて書き表すと，

$$h(s, U) = Sa\left(\frac{k\pi}{\Delta s}s\right) \times Sa\left(\frac{f}{z_0}K_{\Omega_u}U\right) \quad (6.16)$$

となり， s に関する補間と， U に関する補間の積の形で表される．したがって， s, U についてそれぞれ独立に，補間処理を施せばよいことが分かる．第2項の U 軸上の補間は，ハードウェアのテクスチャマッピング機能の補間のオプション（線形補間）をそのまま用いることにした．したがって，プログラム側では，(6.16) 式の第1項の， s 軸上の補間のみを行う．

本来，light field は4次元空間 (s, t, u, v) であるため，プログラム側で実装すべきフィルタ $w(s, t)$ は，下記のとおりである．

$$w(s, t) = Sa\left(\frac{k\pi}{\Delta s}s\right) Sa\left(\frac{k\pi}{\Delta t}t\right) \quad (6.17)$$

本論文では，この理想的な波形に対して，次のような線形近似（一次ホールド）を採用する．

$|ks| < \Delta s$ かつ $|kt| < \Delta t$ のとき，

$$w(s, t) = \left(1 - \left|\frac{ks}{\Delta s}\right|\right) \left(1 - \left|\frac{kt}{\Delta t}\right|\right)$$

それ以外の場合，

$$w(s, t) = 0 \quad (6.18)$$

上記の線形近似による，normal filter，wide-aperture filter，camera-skipped filter の波形を図 6.19 に示す．各格子点は，多視点画像が取得された位置（サンプリング点）を表す．camera-skipped filter では入力画像が間引かれるため，画像合成に使用されないカメラの位置を白抜きの円で示した．定数 A は，各サンプリング点に $w(s, t)$ を畳み込んだときに， (s, t) 面上の任意の点で，ゲインが1になるように調整される．本論文における画像合成は，すべてこれらの近似的な波形で実装されている．

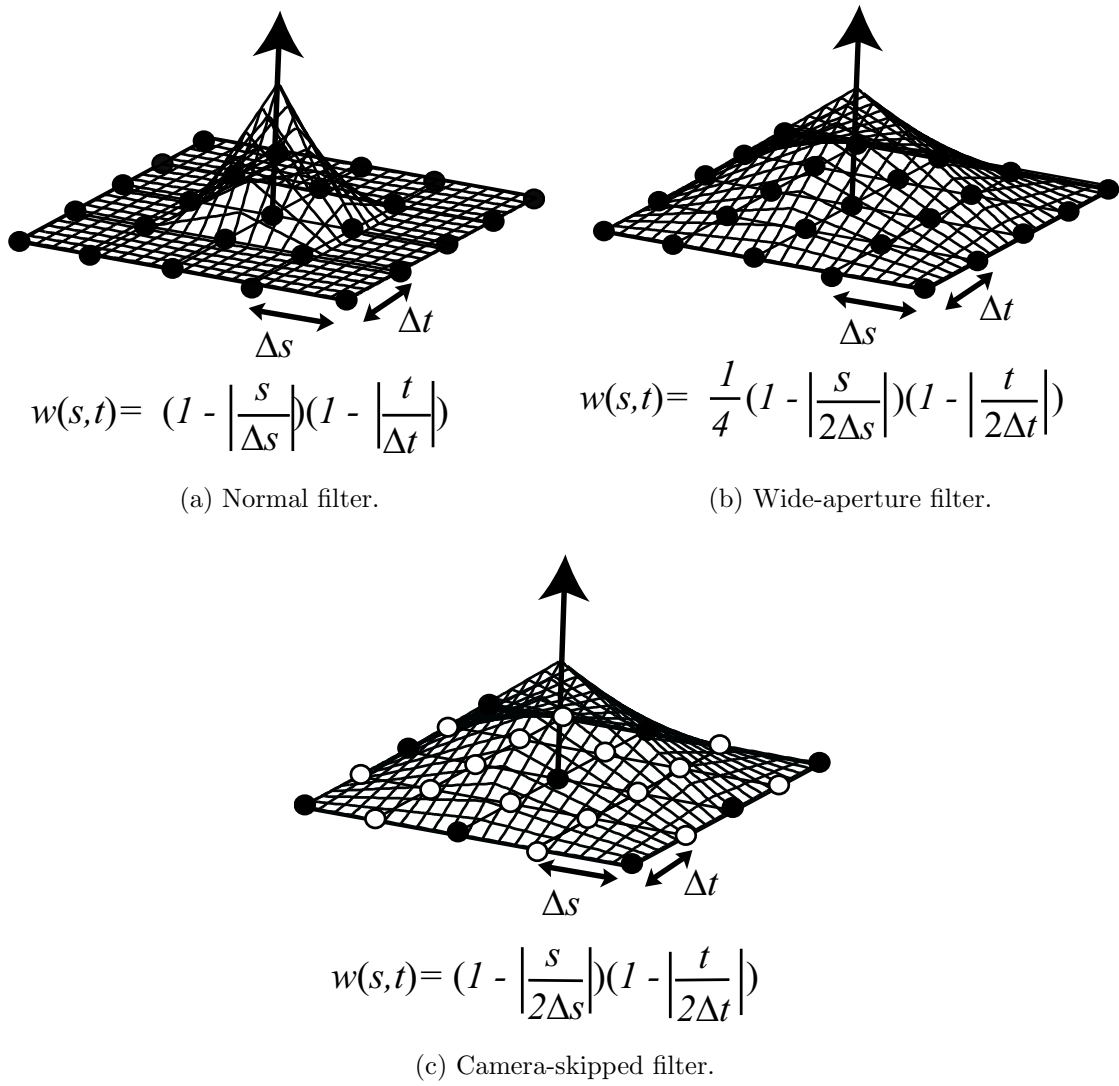


図 6.19: 空間領域における各フィルタの実装 .

第7章

合焦判定に基づく 全焦点自由視点画像合成法の 実システムへの適用

7.1 まえがき

本論文の第5, 6章では, 合焦判定に基づき, 全焦点・自由視点画像を合成する手法を提案し, 静止空間を対象とした実験により, 本手法の有効性を確認した。しかしながら, 提案手法の特徴はオフライン処理を一切必要としない実時間合成にあるため, 動空間に適用されて初めて, その長所が十分に示されることになる。そこで, 本章では, レンズアレイを用いた実時間システム LIFLET (Light Field Live with Thousands of Lenslets)[1, 2] に, 提案手法を適用することを試みる。

従来の LIFLET では, ある視点から見た被写体の奥行きを, いくつかの代表点について求め, それを基に構成された“簡易構造モデル”を用いて画像合成を行っていた。しかし, この手法では, 対象によっては, モデルの解像度が必ずしも充分であるとはいえないため, 満足な合成品質が得られない場合もあった。一方, 合焦判定に基づく全焦点合成法では, 合成画像の各画素に固有の奥行きが割り当てられる。したがって, この合成法を当てはめることにより, 従来より高品質な画像合成が可能になることが期待される。

なお, “合焦判定”法はカメラアレイによる多視点画像を想定して提案された手法であるため, これを LIFLET の撮像系で取得した多視点画像に適用する際, 合成手法を最適化することが, 一つの課題となる。また, 多視点画像の取得と, 画像合成を同時に行うため, 処理の並列化等のシステム全体としてのチューニングも重要なテーマである。

7.2 従来研究

本章では提案システムで用いる, LIFLET の撮像系と“合焦判定”に基づく全焦点画像合成法の概要について説明する。また, レンズアレイ画像を用いた自由視点画像合成について, これまでの研究成果を概観する。

7.2.1 LIFLET の撮像系

LIFLET の撮像系の外見を図 7.1 に示す。撮像系は, 奥行き制御レンズ(凸レンズ), レンズアレイ, 一台のカメラからなる。まず, 被写体空間は, 奥行き制御レンズによって, レンズアレイの入射端面付近に結像される。なお, この像は単に縮小されるだけでなく, 奥行き方向に非線形に歪んだものになる。次に, レンズアレイにより, それぞれのレンズが各位置からみた「被写体空間の像」が, 出射端面付近に結像される。ここで用いるレンズアレイは, 直径約 1mm のマイクロレンズを数千眼, 密に並べたものであり, 個々のレンズが一台のカメラの役割をする [3]。そのため, カメラでレンズアレイを撮像すると, 被写体空間(の像)の多視点画像が取得できる。このように, レンズアレイを用いて取得した多視点画像は, その全体を“IP (Integral Photography) 画像”と呼び, レンズ一眼に相当する領域を“要素画像”と呼ぶ。IP 画像の例を図 7.2 に示す。

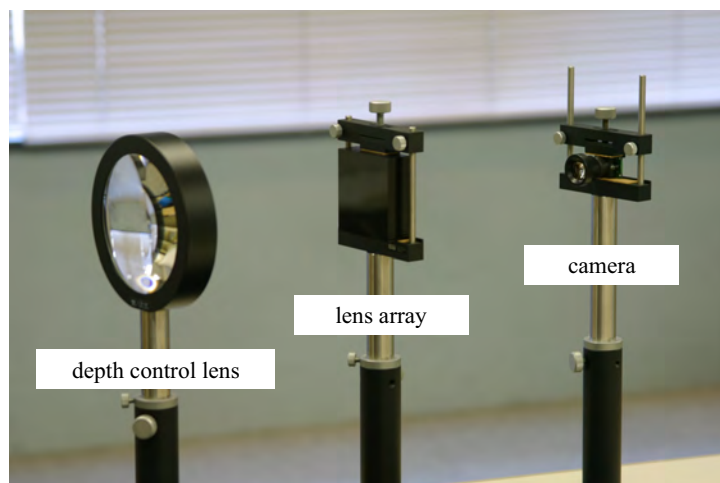


図 7.1: LIFLET の撮像系

7.2.2 合焦判定に基づく全焦点画像合成

第5, 6章では, カメラアレイによる入力を想定し, “合焦” 判定に基づく“全焦点” 画像合成法を提案した. ここでは, この手法について概説する.

提案手法は, light field rendering [4] を拡張したものである. light field とは, 光の満たされた実空間のことであり, この光線データは4次元情報空間である ray space に記録される. 自由視点画像合成の際には, ある視点に到着する各光線の経路を追跡し, ray space から必要な光線を読み出し, 補間などを行う. 被写体の形状が未知の場合, 光線の経路を追跡するにあたって, 三次元空間を focal plane と呼ばれる一枚の平面に近似する. その結果, 合成画像では focal plane の付近の物体のみが鮮明に合成され, focal plane から離れた物体はぼけて合成される [5, 6].

“合焦” 判定に基づく合成法は, 多数枚の focal plane を用いることにより, 上記の問題を解決するものである. まず, focal plane の奥行きを前後に変えながら, 多数枚の画像を合成する. 次に, これらの画像から鮮明な領域を抽出 (“合焦” 判定) し, 一枚の画像に統合することによって, すべての奥行きで鮮明な画像を合成する. なお, 各画像から鮮明な領域を抽出する “合焦” 判定は, 同じ focal plane に対し2種類のモードで画像を合成し, それらの差分を評価することで実現される.

7.2.3 レンズアレイ画像を用いた自由視点画像合成

レンズアレイを用いて取得された画像 (IP 画像) は, 通常, 同様のレンズアレイを組み込んだ立体ディスプレイで表示するために用いられてきた [7]. 一方, 本研究室では, 2000 年度より, レンズアレイで撮像された画像から自由視点画像合成を行う研究を継続している.

まず吉田ら [8, 9, 10] は, 立体ディスプレイを介して IP 画像を見た場合の見えをシミュレートすることにより, 自由視点画像を合成する手法を提案した. しかし, この手法においては, 個々の要素画像が一つの色に見えるため, 解像度が低いという問題点があった. そこで, 吉田らはさ

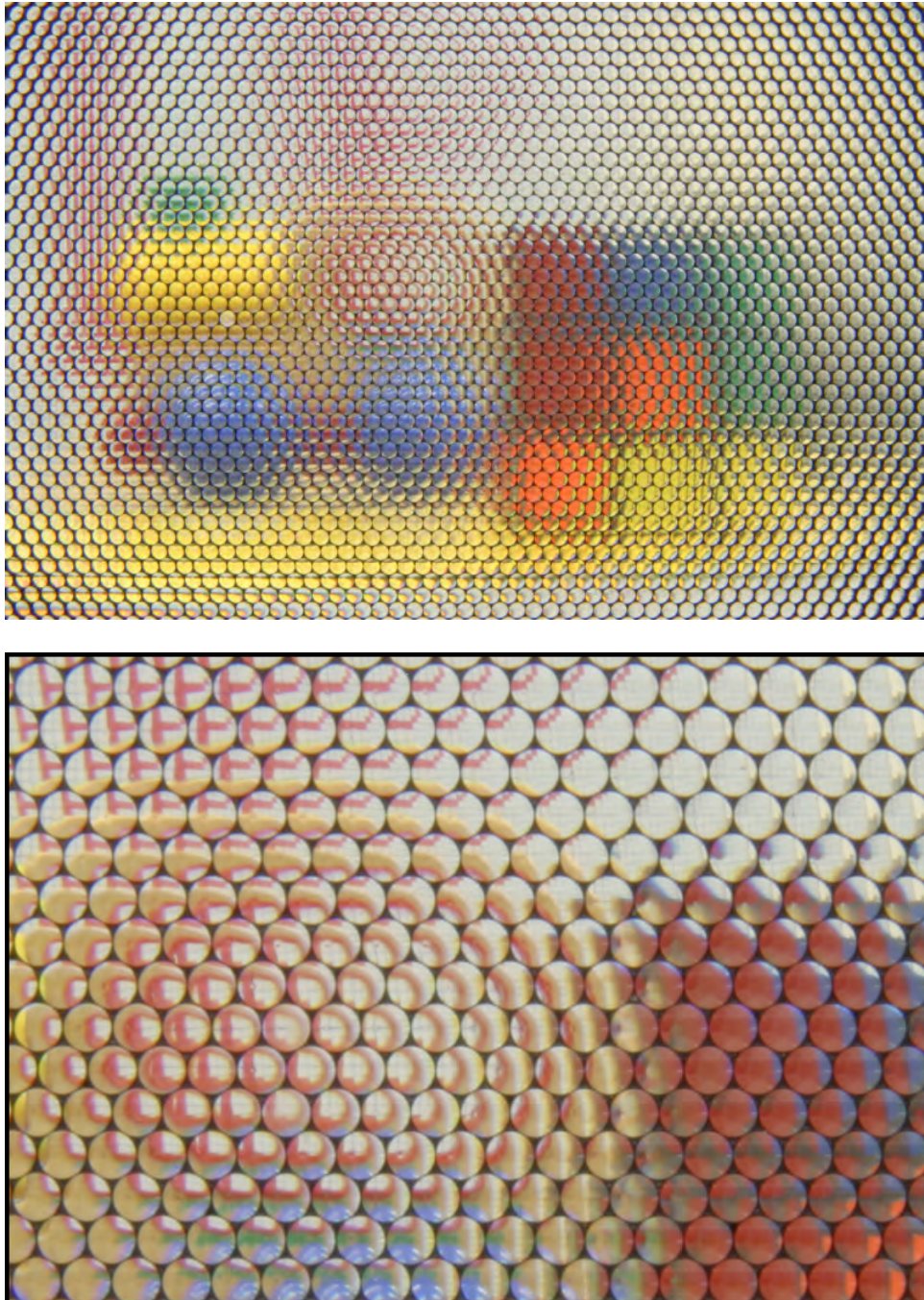


図 7.2: 撮影された IP 画像の例（上）と中央付近の拡大図（下）

らに、奥行き仮定平面を用いた補間を導入することにより、より解像度の高い画像合成を行う手法を開発した。この奥行き仮定平面は、light field rendering における focal plane と等価であるため、合成画像上では、同様の“焦点ぼけ”の効果が観察された。

光田ら [11, 12] は、吉田らの手法を改善して画質を向上させるため、実時間の奥行き推定処理を導入した。光田らの手法では、まず、ステレオマッチング法を利用して、要素画像ごとにひとつの奥行きを推定する。次に、推定された奥行きを頂点とした簡易なポリゴンモデルを構築し、このモデルを仮定した画像合成を行う。このポリゴンモデルは、画像を合成する視点位置に依存するものであり、“視点依存簡易モデル”と呼ばれる。吉田らの手法では、奥行き仮定平面から離れた被写体は不鮮明になる傾向があったが、光田らの手法により、被写体空間全体を鮮明に捉えた画像合成が可能になった。

上記の検討は、いずれも静止画を対象としたものであった。山本ら [1, 2, 13, 14] は、レンズアレイとビデオカメラを組み合わせて、動空間を実時間でレンダリングするシステムを開発した。このシステムは、LIFLET (Light Field Live with Thousands of Lenslets) と命名された。さらに、山本らは、レンズアレイを用いた撮像系における奥行き制御レンズ (depth control lens) の影響について解析し、画像合成の際に、奥行き方向の非線形な変形を元に戻す手法も提案した。

小島ら [15, 16, 17] は、LIFLET の撮像系における周波数特性を理論的に解析し、カメラの倍率や奥行き制御レンズおよび被写体の配置が、撮像される IP 画像の鮮明度に与える影響を明らかにした。この理論により、それまで経験に頼ってきた撮影条件の設定について、定性的傾向を踏まえた見通しを立てることが可能になった。また小島らは LIFLET を拡張し、画像の取得に四台のカメラを用いるシステムの構築し、解像度の向上を実現した。

本章における検討課題は、実時間システム LIFLET に対して、全焦点・自由視点画像合成法を適用することである。“視点依存簡易モデル”とは異なり、この手法では、合成画像の各画素に奥行きが割り当てられるため、より鮮明な画像が合成できることが期待される。また、これまでに蓄積された撮像光学系についての知識も、本検討において必須のものである。

7.3 レンズアレイ方式自由視点画像合成における“合焦判定”法の適用

本章では、レンズアレイを用いて多視点画像を取得し、“合焦”判定法を用いて自由視点画像を合成するシステムを提案する。なお、LIFLET の撮像系では、カメラアレイとは異なり、奥行き制御レンズが利用されている。そのため、“合焦”判定法の適用の際に、撮像光学系に合わせて、合成アルゴリズムを変更する必要がある。奥行き制御レンズは、被写体空間の適切な奥行き範囲を鮮明に撮影するために使用するものであるが [16]、前述のように、その像は単に縮小されるだけでなく、奥行き方向に非線形に歪んだものとなる。取得された IP 画像は、レンズによって結ばれた像の多視点画像である。そこで、画像合成の際に特に下記の二点に注意する。

- 奥行き制御レンズによる光線の屈折を考慮した light field rendering
- focal plane の設定の最適化

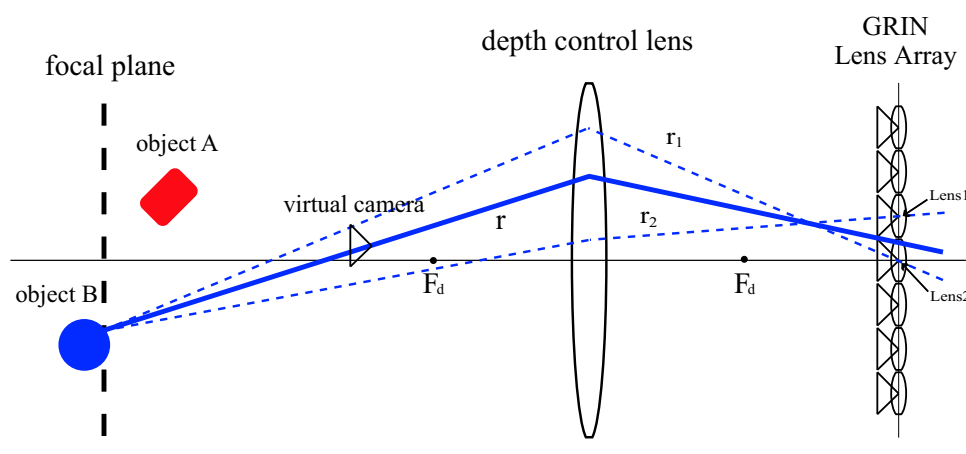


図 7.3: LIFLET 光学系を用いた light field rendering

それぞれの詳細を以下で説明する．

7.3.1 奥行き制御レンズを考慮した light field rendering

吉田らの手法 [9] では図 7.1 のような撮像系で IP 画像を取得し、一枚の focal plane を用いて自由視点画像を合成する手法が提案されている．これは“合焦”判定に基づく画像合成法で用いられる light field rendering と等価な手法である．しかし、この手法では奥行き制御レンズの影響を考慮していないため、あくまでも、レンズによって結像された空間をレンダリングしていたにすぎない．一方、山本らの手法 [1, 2] では、光線の経路を追跡する際に、奥行き制御レンズによる屈折を考慮し、結像される前の被写体空間をより正確にレンダリングする手法が用いられている．そこで本章では、文献 [9] の手法に対して、文献 [1, 2] のような屈折を考慮したアプローチを導入し、light field rendering を行うことにする．

図 7.3 を用いて画像合成の手法を説明する．図中の virtual camera は、画像を合成する視点位置を表す．所望の画像は、virtual camera の中心を通過するすべての光線の色を求めることによって合成される．例として、図中に実線で示した光線 r を合成する場合を考える．まず、focal plane を定義し、すべての光線がこの面上から発していると仮定する．これは、三次元の被写体空間を二次元の focal plane によって近似することを意味する．次に、この光線が、レンズアレイを通過する位置を計算し、近傍のレンズ (Lens1, Lens2) を求める．続いて、光線 r に対応し、Lens1, Lens2 によって結像されている光線、 r_1 , r_2 を求める．ここで述べる対応とは、focal plane 上で同一の点を通過することを意味する．光線 r の色は、光線 r_1 と r_2 の色を線形に重み付けすることによって求められる．ここで重要なのは、光線の経路を追跡する際に、depth control lens による屈折を考慮して計算を行っていることである．

第 5 章では、2 つのモードを用いて画像を合成し、それらの比較により、“合焦領域”を抽出する手法を提案した．ここでは、図 7.4 を用いて、レンズアレイを用いた撮像系でも同様の原理が成り立つことを示す．図 7.4(a) は、base mode であり、すべての入力画像を利用して画像を合成

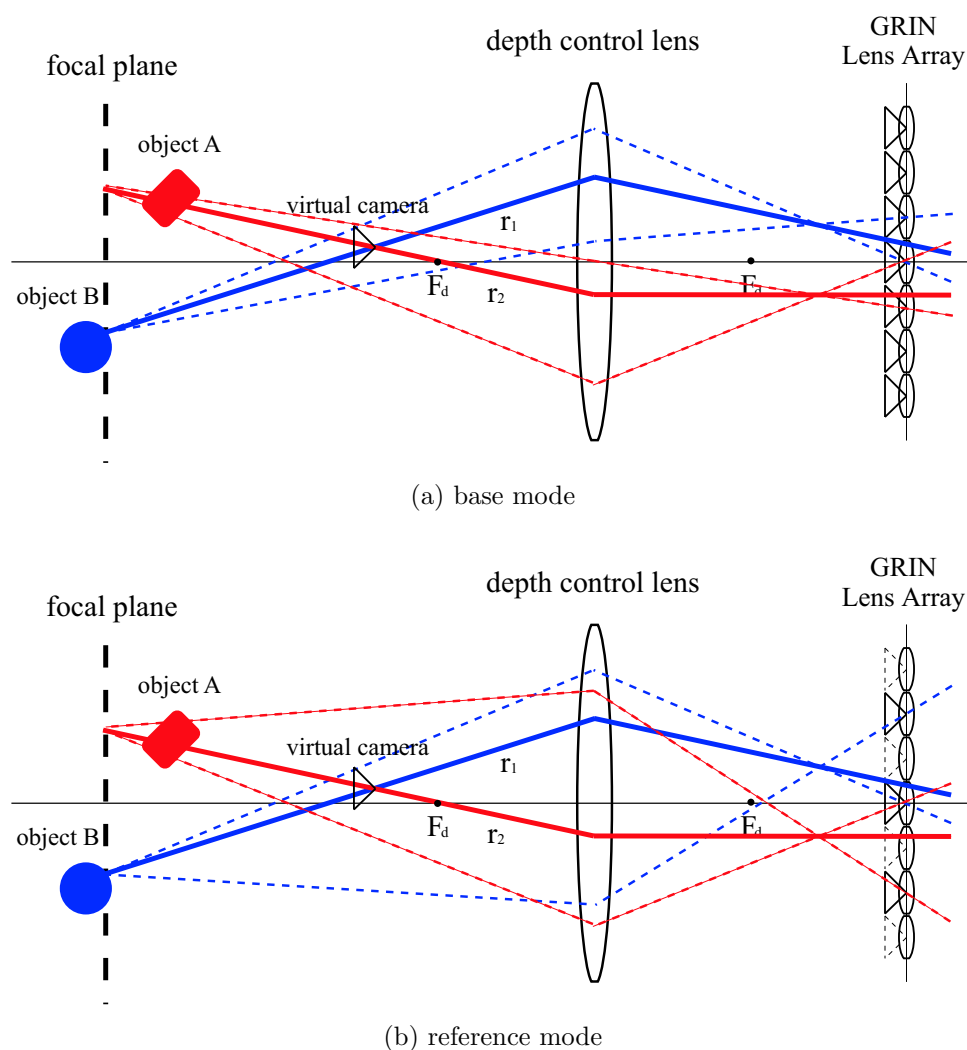


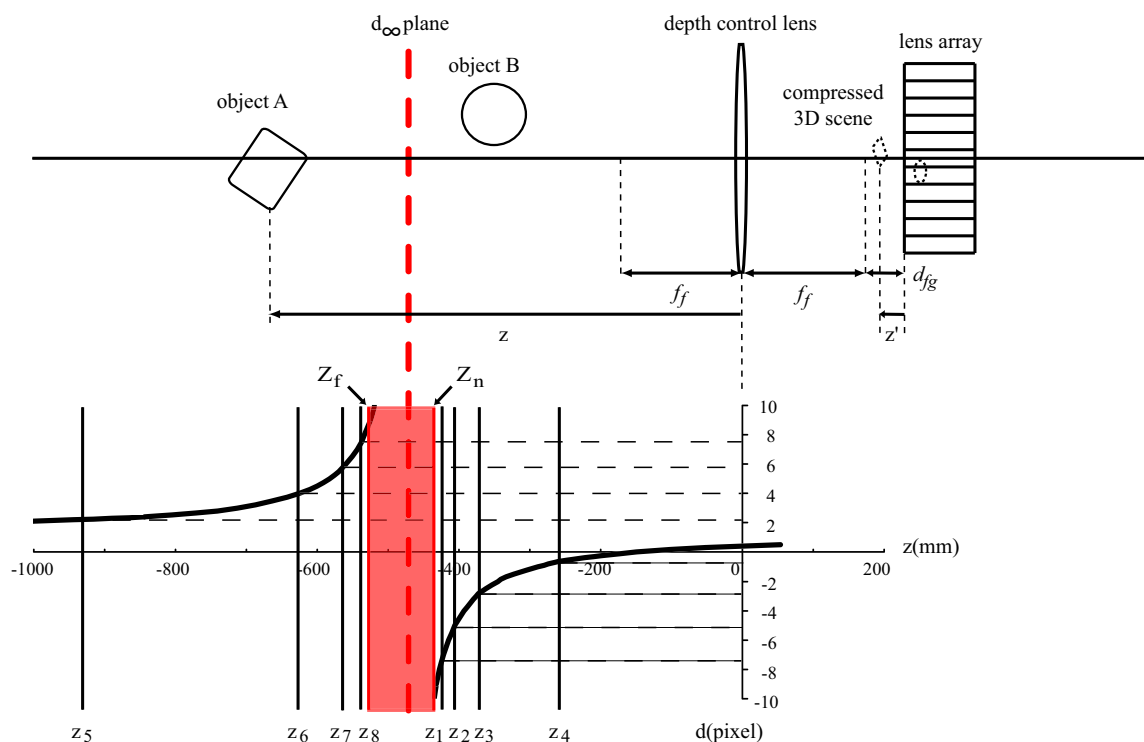
図 7.4: LIFLET 光学系における“合焦判定”。

する．一方，図 7.4(b) に示す reference mode では，入力画像を一つおきに間引いて利用する．利用されない画像を破線で示す．双方のモードにおいて，focal plane は Object B の位置に配置されている．したがって，Object B には“焦点”が合うが，Object A は“焦点”から外れる．

ここで，物体の反射特性がランバートであり，かつオクルージョンの問題が無視できると仮定する．物体の表面が focal plane 上にある場合，両方のモードで合成される光線の色は同じである（例えば，光線 r_1 ）．これは，どちらのモードでも，被写体表面上の同一の点から発する光線が合成に利用され，ランバート仮定により，これらの光線の色は同じだからである．一方，物体が focal plane 上でない場合，両者において合成される光線の色には違いが生じる（例えば，光線 r_2 ）．これは，両モードにおいて，被写体空間の異なる点から発する光線が合成に利用されるためである．したがって，この二つのモードによる合成画像の違いを評価することにより，“焦点が合っている”領域を抽出することが可能である．この結論は，第 5 章の議論によって得られた結

表 7.1: 各種パラメータ

z	distance between objects and depth control lens
z'	distance between image of objects and lens array
f_f	focal length of depth control lens
f_g	focal length of lens array
d_{fg}	distance between depth control lens focus and lens array

図 7.5: レンズアレイを用いた撮像系（上）と focal plane (z_1 – z_8) の設定の最適化（下）

論と同じである．以上の議論により，depth control lens を用いた LIFLET の撮像系においても，同様の“合焦判定”法が適用可能であることが原理的に確かめられた．

7.3.2 focal plane の設定の最適化

“合焦”判定による合成法では，まず，多数枚の focal plane を用いて light field rendering により画像合成を行う．これを LIFLET の撮像系で取得した多視点画像に適用する場合には，focal plane の設定に注意が必要である．

図 7.5（上）の撮像系の各パラメータの説明を表 7.1 に示す．これらのパラメータを用いて，被写体の奥行き z と隣り合う要素画像間の視差 d の関係式を導くことができる．まず，被写体空間

の像の奥行き z' と視差 d は、次式に示すように、反比例の関係にある。

$$z' = \frac{n_p f_g}{d} \quad (7.1)$$

なお、 n_p は要素画像 1 辺の画素数を表す。一方、ニュートンの結像公式より、

$$(z - f_f)(d_{fg} - z') = f_f^2 \quad (7.2)$$

が成り立つ (7.1) 式を (7.2) 式に代入し、 z を d の関数と考えたと次式を導くことができる。

$$z(d) = \frac{f_f^2 n_p f_g}{d_{fg}^2} \cdot \frac{1}{d - \frac{n_p f_g}{d_{fg}}} + \frac{f_f^2}{d_{fg}} + f_f \quad (7.3)$$

(7.3) 式を基に、 z と d の関係をプロットしたグラフを図 7.5 (下、太い実線) に示す。グラフで、 $z = 0$ は、図 7.5 (上) の奥行き制御レンズの位置と一致させている。図 7.5 からわかるように、視差 d は正の値だけでなく負の値も持ち、その値が無限大となる奥行き面が存在する。これは奥行き制御レンズを用いることに起因するものであり、本章では、この視差が無限大になる奥行き面を“視差無限大平面”と呼ぶことにする。視差無限大平面の付近は、多視点画像にならないため¹、この付近 ($Z_f < z < Z_n$ とする) には被写体を置かないことにする。

focal plane を N 個配置する場合を考える。まず、視差無限大平面の前後に配置する focal plane の枚数 (N_n, N_f) を次式に従い定める。

$$N_n = \left\lceil \frac{D_n}{D_n + D_f} \cdot N \right\rceil \quad (7.4)$$

$$N_f = N - N_n \quad (7.5)$$

なお、 $[x]$ は x を超えない最大の整数を意味する。 D_n および D_f は視差無限大平面の前後の視差距離であり、次式で求められる。

$$D_n = |d(Z_n) - d(0)| \quad (7.6)$$

$$D_f = |d(-\infty) - d(Z_f)| \quad (7.7)$$

なお、 $d(Z_n)$ は $z = Z_n$ に対応する視差である。この時、focal plane の奥行き z_n は、Chai らの plenoptic sampling 理論 [18] に従い、視差無限大平面前後のそれぞれの視差空間を等分するように設定する。

$$z_n = z(d_n) \quad (7.8)$$

$$d_n = \begin{cases} d(Z_n) + D_n \cdot \frac{n - 1/2}{N_n} & (n \leq N_n) \\ d(\infty) + D_f \cdot \frac{(n - N_n) - 1/2}{N_f} & (n > N_n) \end{cases} \quad (7.9)$$

インデックスの n は、視差 d_n が単調増加するように定めている。 $N = 8$ の時の設定例を図 7.5 (下) に示す。 z_1 から z_8 は上記の方法で定められた focal plane の奥行きである。

¹視差無限大平面付近のある 1 点は、複数の要素画像に同時に撮影されない

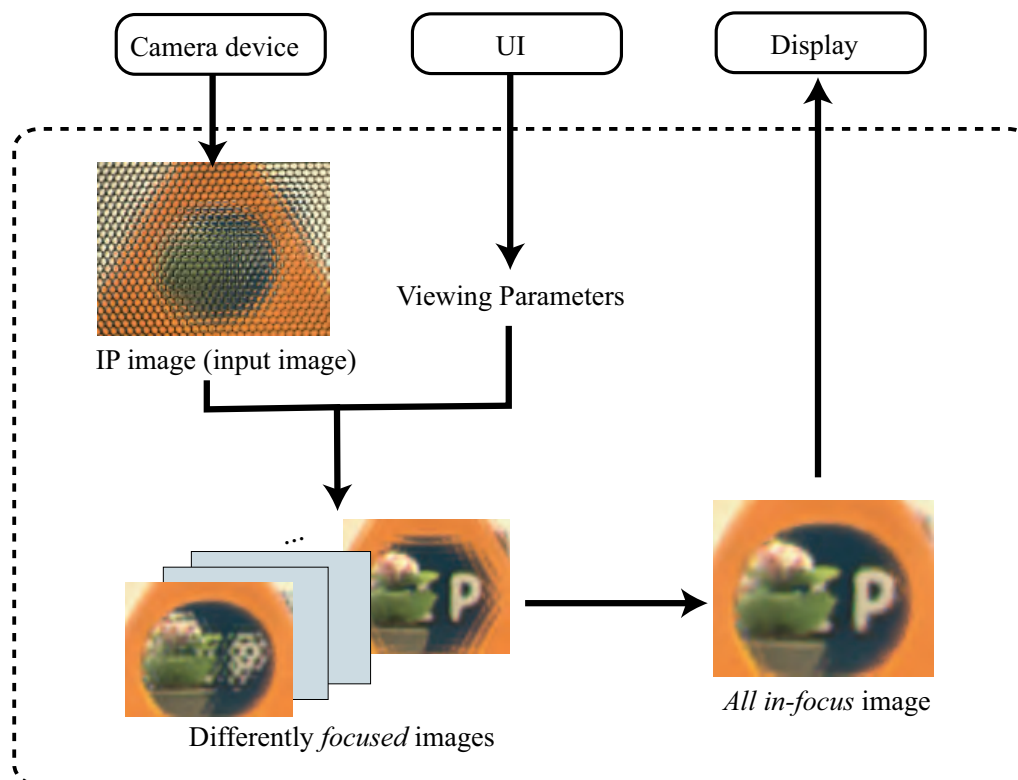


図 7.6: 画像合成の過程

7.4 実時間合成のための実装および実験

7.4.1 実時間合成

図 7.6 に処理のおおまかな流れを示す。本システムは 1 台の PC 上で動作する。システムには、カメラデバイスを通して IP 画像が、UI を通してレンダリングのパラメータが逐次入力される。本実験では、カメラとして、ポイントグレイリサーチ社の Dragonfly デジタルカメラ（1024×768 画素，15 fps，IEEE1394 インターフェイス）を用いた。またレンダリングパラメータは、マウスの操作により直感的に制御できる設計とした。システム内部では、まず、与えられた視点位置に対して、“合焦”距離を変えながら多数の画像を合成する。次にこれらの画像の“焦点”が合う画素を選択的に統合することで、“全焦点”画像を合成し、ディスプレイに表示する。実験の様子を、図 7.7 に示す。ここでは、被写体として 4 つの物体：手前から、橙色の三角形、花の模型、積み木（“IP”と書いてある）、背景面（“LF”と書いてある）を配置した場合を示している。

画像合成の詳細な過程は下記の通りである。

1. 与えられた視点位置において、“合焦”距離（focal plane の奥行き）を変えながら、多数枚の画像を合成する。このとき、それぞれの“合焦”距離において、base mode および reference mode で各一枚の画像を合成し、両者の差分を取る。
2. それぞれの“合焦”距離において、差分画像を平滑化する。



図 7.7: 実験風景

3. 平滑化された差分画像群を用いて、画素ごとに奥行き之最適値を探索し、奥行き画像を得る。さらに得られた奥行き画像を平滑化する。
4. 奥行き画像を元に、base mode による合成画像群から、最も“焦点”が合う画素を選択し、“全焦点”画像を合成し、画面に表示する。
5. カメラから IP 画像を取得し、1. に戻る。

上記の各処理における処理時間の内訳を計測した結果を、図 7.8 に示す。使用した PC には、Xeon 3.2 GHz の CPU×2 基、2.0 GByte のメインメモリが搭載され、ビデオカードには、nVIDIA GeForce FX 5900XT のグラフィックスプロセッサ、128 MByte のビデオメモリが搭載されている。なお、ここでは、focal plane の枚数は 8、画面サイズは 400×300 画素とした。逐次処理（単スレッドによる処理）の場合を図 7.8 (a) に、並列化（マルチスレッド化）した場合のイメージを図 7.8 (b) に示す。グラフの凡例は図 7.8 (c) の通りである。(b) については、並列処理の進行状況を直接計測することが難しかったので、(a) のデータに基づいて描いたイメージ図とした。

図 7.9 に、focal plane の枚数を変えながら全体の処理時間を計測した結果を示す。こちらのグラフにおいては、並列化した場合についても実測データである。並列化によって、効率的にスループットを改善できていることがわかる。必要なレイヤの数はシーンの持つ奥行きに依存するが、本実験においては 8 枚程度で満足な画質が得られている。この条件下では並列化した場合の処理速度は 10.9 fps であり、動空間のインタラクティブなレンダリングに十分な速度が得られた。

動空間を対象にした自由視点画像合成の様子を図 7.10 に示す。左図では水の流れる様子や、水によって光が屈折される様子が、右図では穴の開いた物体によって生じるオクルージョンが正しく再現されていることが確認できる。

7.4.2 従来手法との画質の比較

ここでは、従来の LIFLET の合成手法を用いた場合 [1, 2] と“全焦点”合成手法（提案手法）を用いた場合の画質比較を行う。先の実時間合成実験では、Dragonfly デジタルカメラ（1024×768

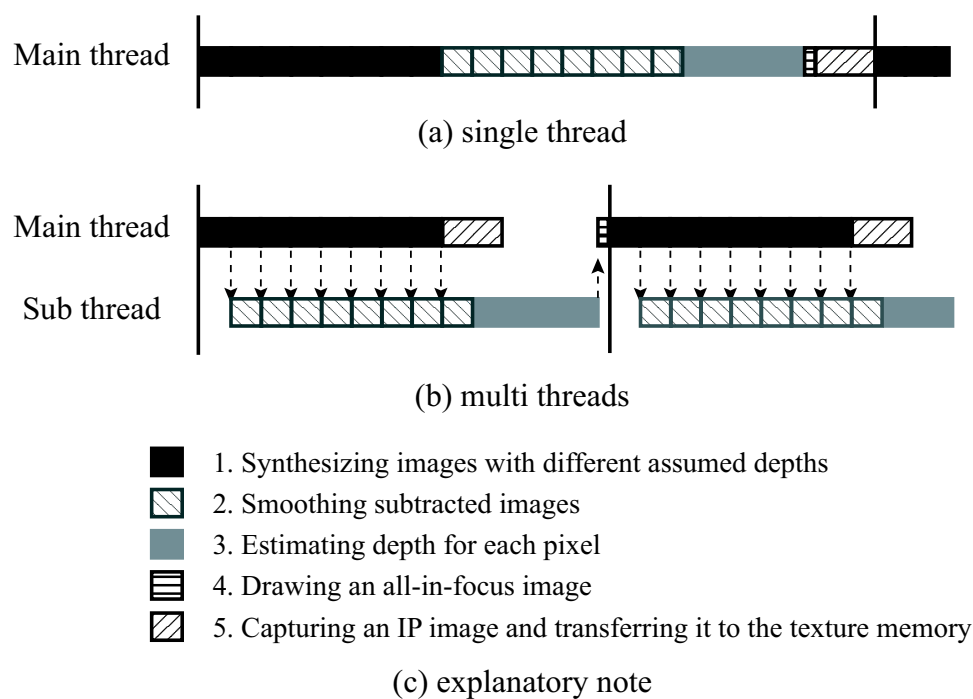


図 7.8: マルチスレッドを用いた実装

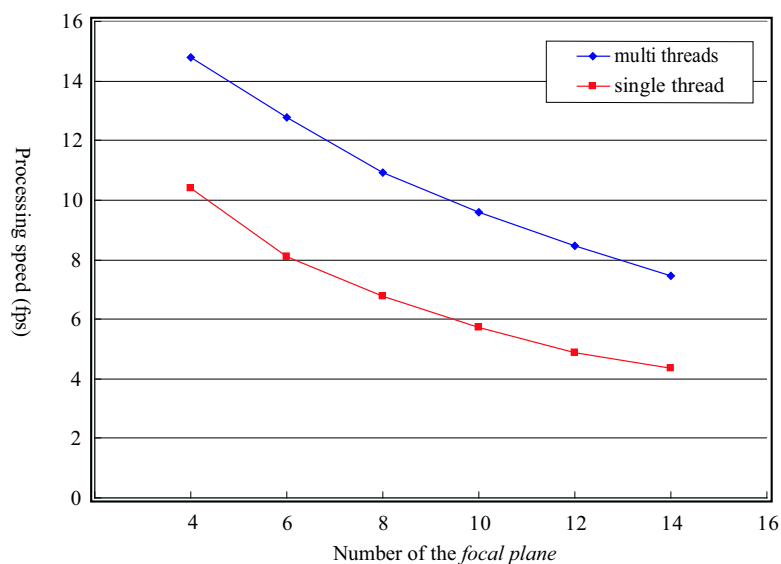


図 7.9: 処理速度

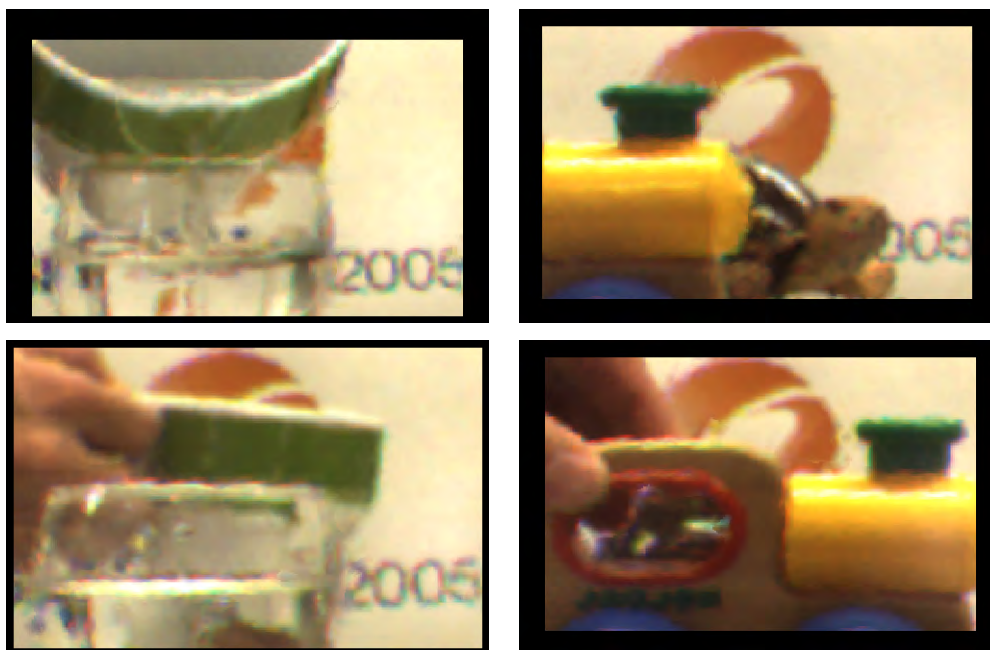


図 7.10: 動空間に対象にした合成の様子

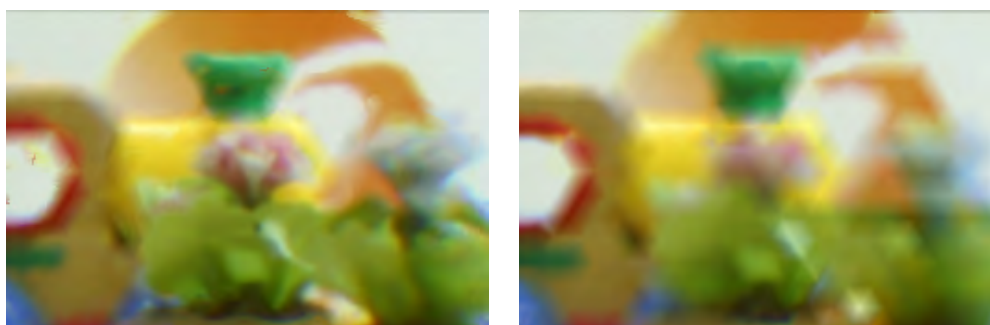


図 7.11: 提案手法（左）と従来手法（右）の画質の比較

画素)を使用した。この解像度においても、両者の画質の差を確認できるが、さらに高解像度のカメラを用いることで、その差はいっそう明確になる。そこで、高解像度カメラ (Canon EOS D60: 3076×2048 画素) を用いた実験もあわせて行った。ここでは、要素画像の直径 $n_p=130$ 画素とした。なお、カメラのインターフェイスの制約上、撮影したのは静止 IP 画像である。

図 7.11 に、合成画像の例を示す。左が提案手法、右が従来の LIFLET による合成結果である。提案手法によって、より鮮明な画像が合成できていることがわかる。この品質の違いは、両者において推定されるモデルの解像度の差に起因する。従来の LIFLET では、要素画像一つにつき一つの奥行きを持つ、“視点依存簡易モデル” が用いられるのに対し、提案手法では、合成画像の個々の画素に固有の奥行きが与えられる。したがって、入力画像の解像度が高くなるほど、提案手法を導入することのメリットが顕著になるといえる。



図 7.12: 本研究室において開発しているカメラアレイ。

7.5 むすび

本章では、本論文の第5, 6章で提案した“合焦判定”に基づく“全焦点”自由視点画像合成法を、レンズアレイを用いた実時間システム LIFLET に適用した結果について述べた。提案手法は、もとともカメラアレイによる入力を想定したものであるが、撮像系にあわせたチューニングを行うことで、レンズアレイを用いたシステムにも適用可能であることが確認できた。具体的には、奥行き制御レンズの影響を考慮して、light field rendering のアルゴリズム、および focal plane の設定方法に調整を加えた。また、一台の PC を用いた実験により、動空間の実時間レンダリングが可能であることを確認した。さらに、“視点依存簡易モデル”を用いていた従来のレンダリング手法と比較して、合成画像の品質が向上することも確かめられた。

なお、現段階の合成画像にはまだ“合焦判定”に失敗する部分があるため、今後、その原因を明らかにし、さらに画質を改善する必要がある。また、実カメラアレイ入力に対する提案手法の適用についても検討する。文献 [19, 20, 21, 22] では、数十台–百台程度のカメラを用いたシステムの実時間動作が報告されており、本研究室でも、128 台のネットワークカメラを用いたカメラアレイシステム構築に関して、基礎検討を始めている（図 7.12）。

参考文献

- [1] T. Yamamoto and T. Naemura, “Real-Time Capturing and Interactive Synthesis of 3D Scenes Using Integral Photography,” IS&T/SPIE Electronic Imaging 2004, Stereoscopic Displays and Applications XV, 5291A-16 (2004).
- [2] T. Yamamoto, M. Kojima, and T. Naemura, “LIFLET: Light Field Live with Thousands

- of Lenslets”, ACM SIGGRAPH 2004 Emerging Technologies, etech.0130 (2004).
- [3] 洗井 淳, 岡野文男, 星野春男, 湯山一郎, “屈折率分布レンズを用いたインテグラルフォトグラフィ撮像方式の基礎検討”, 映情学誌, Vol. 51, No. 11, pp. 1927–1934 (1997.11).
- [4] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH, pp. 31–42 (1996.8).
- [5] A. Isaksen, M. Leonard, and S.J. Gortler, “Dynamically Reparameterized Light Fields”, Technical Report MIT-LCS-TR-778 (1999).
- [6] A. Isaksen, L. McMillan, and S.J. Gortler, “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH, pp. 297–306 (2000).
- [7] M.G. Lippmann, “Epreuves reversibles donnant la sensation du relief,” J. de Phys., Vol. 7, No. 4, pp. 821–825 (1908).
- [8] 吉田達哉, 苗村 健, 原島 博, “インテグラルフォトグラフィを用いたインタラクティブ3次元CGの合成”, 3次元画像コンファレンス2000, pp. 39–42 (2000).
- [9] T. Naemura, T. Yoshida and H. Harashima, “3-D Computer Graphics Based on Integral Photography”, Opt. Express, Vol. 8, No. 4, pp. 255–262 (2001).
- [10] 吉田達哉, 苗村 健, 原島 博, “インテグラルフォトグラフィを用いた3次元CGの合成”, 映情学誌, Vol. 55, No. 3, pp. 474–478 (2001).
- [11] 光田智史, 山本智幸, 高橋桂太, 苗村 健, 原島 博, “実時間奥行き推定を用いたインテグラルフォトグラフィからの自由視点画像合成”, 3次元画像コンファレンス2003, pp. 13–16 (2003).
- [12] S. Mitsuda, T. Yamamoto, K. Takahashi, T. Naemura, and H. Harashima, “Interactive View Synthesis from Integral Photography Using Estimated Depth Information,” Proceedings of SPIE Vol. 5243 Three-Dimensional TV, Video, and Display II, pp. 116 – 124 (2003).
- [13] 山本智幸, 苗村 健, “レンズアレイ方式IBRシステムにおける実時間処理の検討”, FIT2003 情報科学技術フォーラム 一般講演論文集, 第3分冊, J-082, pp. 365–366 (2003).
- [14] 山本智幸, 苗村 健, “屈折率分布レンズ方式IBRシステムにおける奥行き圧縮効果の除去”, 信学技報 IE2003-105 (2003.11)
- [15] Masaru Kojima and Takeshi Naemura: “McLiflet: multiple cameras for light field live with thousands of lenslets,” Proceedings of SPIE Vol. 5664, Stereoscopic Displays and Virtual Reality Systems XII, pp. 495–505 (2005).

- [16] 小島 将, 苗村 健, “レンズアレイを用いた自由視点画像合成システムにおける撮像光学系の空間周波数特性”, 映情学誌, **59**, 4, pp. 569–580 (2005.4).
- [17] Masaru Kojima and Takeshi Naemura: “McLiflet: multiple cameras for light field live with thousands of lenslets,” accepted to SPIE Stereoscopic Displays and Applications XVI, 5664A-78 (2005).
- [18] J.-X. Chai, X. Tong, S.-C. Chany, and H. -Y. Shum, “Plenoptic Sampling,” Proc. ACM SIGGRAPH, pp. 307–318 (2000).
- [19] T. Naemura, J. Tago, and H. Harashima, “Real-Time Video-Based Modeling and Rendering of 3D Scenes”, IEEE Computer Graphics and Applications, Vol. 22, No. 2, pp. 66–73 (2002.3).
- [20] J.C. Yang, M. Everett, C. Buehler, and L. McMillan, “A Real-Time Distributed Light Field Camera”, Proc. 13th Eurographics Workshop on Rendering, pp. 77–85 (2002.6).
- [21] B. Wilburn, M. Smulski, K. Lee, and M.A. Horowitz, “The Light Field Video Camera”, Proc. SPIE Vol. 4674, Media Processors 2002, pp. 29–36 (2002).
- [22] C. Zhang and T. Chen, “A Self-Reconfigurable Camera Array,” Proc. Eurographics Symposium on Rendering, pp. 243–254 (2004).

第8章

多次元信号処理から見た 合焦判定法と多眼ステレオ法

8.1 まえがき

多視点画像を元に自由視点画像を合成する手法は image-based rendering と呼ばれ、三次元空間を写實的に表現する有力な手段として注目されている [1]。Image-based rendering では、画像に加えて被写体空間の幾何学モデルを用いることが多く、いかにして被写体の形状を得るか、というのは重要な課題である。

本章では、二次元アレイ状に構成された多視点カメラによる画像群のみを入力とし（カメラパラメータは既知とする）、被写体形状については事前の知識がない場合を考える。また、被写体空間を、カメラアレイ面に平行な複数のレイヤ群によって近似して扱うことにする。レイヤモデルを用いて画像合成を行う場合、合成画像の各画素が、どのレイヤに属するのかを判定する必要がある。これは一種の形状推定である。本論文の第5, 6章では、この課題に対する新しいアプローチとして、“合焦”判定という新たな手法を導入した。一方で、上記のような形状推定問題を解く際に最も一般的に用いられるのは、ステレオ法 [3, 4] である。これには、本論文の第4章で提案した手法も含まれる。

ここまでの議論においては、“合焦”判定法とステレオ法の間を定性的に論じてきたが、定量的・数学的な関係については必ずしも明確ではなかった。本章の目的は、“合焦判定”法と、ステレオ法の数学的な関係を明確にすることである。空間・周波数の双方での解析により、基本的な多眼ステレオ法（オクルージョン処理や領域分割を伴わない手法）が、“合焦判定”法の特殊なケースとして位置づけられることを結論として示す。

8.2 背景

本章では、筆者らが第5, 6章で提案した手法の概略を述べる。提案手法は、light field rendering 法 [5, 6] と独自の“合焦判定”法を組みあわせることにより、レイヤモデルを用いた自由視点画像合成を実現するものである。

まず、light field rendering 法について述べる。この手法では、密に取得された多視点画像データを、光線空間上のデータ (light field data) とみなす。画像合成の際には、所望の視点位置に到達するすべての光線の色を求めるが、その際に、カメラ間を通過する光線については、適切に補間することが必要である。被写体の形状が未知である場合、この補間の際には、被写体空間が一枚の平面 (focal plane) 上にある、という仮定を置く。この仮定は、合成画像上で、“焦点ぼけ”と類似した現象を引き起こす¹[6]。すなわち、focal plane 付近の被写体は鮮明に合成されるが、focal plane から離れた物体には、ぼけや二重像が生じる。

第5, 6章における提案手法では、light field rendering 法を拡張し、全体が鮮明な合成画像（“全焦点”画像）を得ることを目的とする。与えられた視点位置に対して、次の処理を行う。

1. light field rendering 法を用いて、focal plane の位置を変化させながら、複数の画像を合成する。

¹ここでは、入力画像そのものには光学的なぼけがないと仮定する

2. それらの画像の“合焦”領域（鮮明に合成されている領域）を検出し，1 枚の画像に統合する．

(1) において合成される画像の枚数は，被写体空間を近似するレイヤの枚数に対応している．(2) において鮮明な領域を検出する処理を，筆者らは“合焦判定”と呼んでいる．これは，合成画像の各画素に，最も適切なレイヤを割り当てる処理であり，一種の形状推定である．

本章では，第 6 章における camera-skipped filter の考え方を拡張する．拡張された“合焦判定”法は，第 6 章で明確に定義された手法とは異なるものの，共通の枠組みの中に位置づけられる．さらに本章では，改めて定義された“合焦判定”法と，基本的なステレオマッチング法が，数学的に密接な関係にあることを示す．

8.3 light field rendering の周波数解析

本章では，“合焦判定”法の理論的基盤として，light field rendering 法の周波数領域での解釈について述べる．

8.3.1 問題設定

入力として用いる多視点画像は，2 次元格子状に平行かつ等間隔に配置した多数のカメラによって取得されるものとする．カメラを配置した平面（カメラ面）上の位置を (s, t) ，点 (s, t) にあるカメラの画素位置を (u, v) で表す．すると，カメラ面を通過する任意の光線は，4 つのパラメータの組 (s, t, u, v) によって一意に表すことができる．これを，4 次元の light field (s, t, u, v) として定義する．原理的には，light field は連続空間として定義されるが，実際の入力データは，有限の解像度を持つ有限個のカメラによって取得されるため，離散的なサンプル値の集合として与えられる．

任意視点画像を合成する問題は，離散的にサンプリングされた light field の補間問題として捉えられる．Chai ら [2] は，多次元信号の周波数解析という立場から，画像合成を低域フィルタによる再構成としてモデル化している．本章の議論は，これらの解析を発展させたものである．説明を簡単にするため，このあとの議論は，4 次元の light field のうち，2 次元の部分空間 (s, u) に限定する． (s, u) の周波数空間を (ω_s, ω_u) と表記する．

8.3.2 周波数解析

light field の信号特性を扱いやすいものとするため，対象とするシーンについて，次の 2 点を仮定する．

1. 被写体の反射特性はランバートとして扱える．
2. オクルージョンの影響が無視できる．

この条件下においては，信号のスペクトラムは，図 8.1 の影をつけた領域に制限される（導出の過程などの詳細は文献 [2] を参照されたい）．ここでは，対象とする空間の奥行きが $z_{\min} \leq Z \leq z_{\max}$

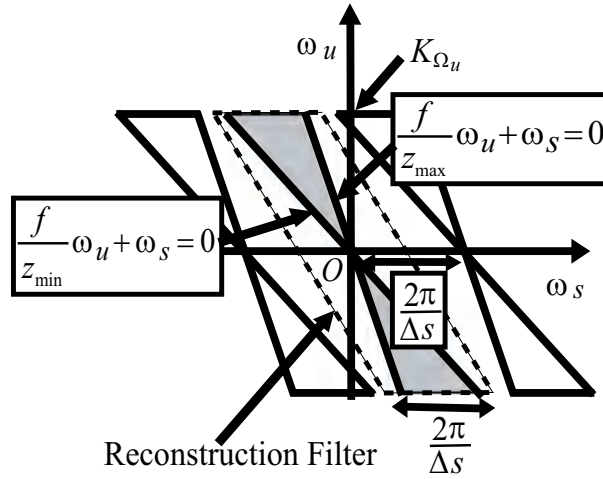


図 8.1: light field の周波数解析 .

で定められるとする． f はカメラの焦点距離を表す． K_{Ω_u} は， u 軸方向の周波数の最大値を表し，シーンの模様の複雑さやカメラおよび合成画像の解像度によって決まる．

Light field は一定のカメラ間隔 Δs で離散的にサンプリングされるため，周波数領域では ω_s 軸の方向に $2\pi/\Delta s$ の周期で原信号の複製成分が繰り返される²．図 8.1 では，左右の隣接する複製成分の輪郭を図示した．複製成分が元の信号と重なり合わないためには，下記の条件が満たされる必要がある．

$$\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \leq \frac{2\pi}{K_{\Omega_u} f \Delta s} \quad (8.1)$$

画像合成の際には，離散的にサンプリングされた光線情報を補間する必要がある．これは，低域フィルタによって，連続信号を再構成することと等価である．Chai ら [2] は，図 8.1 に示すような，幅 $2\pi/\Delta s$ の平行四辺形状の低域フィルタを提案した．このフィルタによって，スペクトラムは，平行四辺形の内部の領域に帯域制限される．このとき平行四辺形の傾きは，画像合成の際に focal plane を仮定する奥行き z_f に対応する（傾き： $-z_f/f$ ）． z_f の最適値 z_{opt} は下記のように定められる．

$$\frac{1}{z_{opt}} = \frac{1}{2} \left(\frac{1}{z_{\min}} + \frac{1}{z_{\max}} \right) \quad (8.2)$$

8.3.3 再構成フィルタと焦点との関係

前節の議論を元に，light field rendering における“焦点”の現象を説明することができる [6]．多くの場合，カメラ間隔 Δs が大きいいため，対象とするシーン全体に対して (8.1) 式を満たすのは困難である．しかし，シーン中の比較的狭い範囲に注目すると，奥行きの変化の幅（ $(1/z_{\min} - 1/z_{\max})$ に対応）が十分に小さく，(8.1) 式が満たされる領域があると考えられる．このような領域を対

²画素間隔を Δu としたとき， ω_u 軸の方向にも $2\pi/\Delta u$ の周期での繰り返しが発生する．

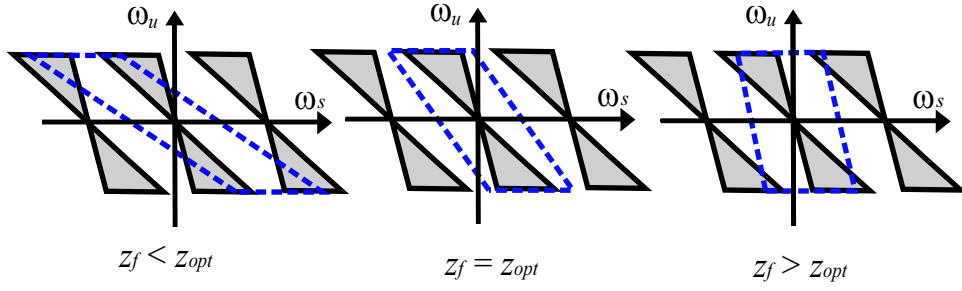


図 8.2: 仮定される奥行きとフィルタの形状 .

象として、ある奥行きを仮定したときの、再構成フィルタの様子を図 8.2 に示す． $z_f = z_{opt}$ のときには、光線信号は、過不足なく適切に再構成される（“焦点” が合う）．しかし、 $z_f < z_{opt}$ や $z_f > z_{opt}$ のときには、再構成において、原信号の高周波成分が漏れたり、近傍の折り返し成分が含まれたりする場合があることが分かる（“焦点” から外れる）．これらは、合成画像において、ぼけ（高周波成分の漏れに対応）や二重像（折り返し成分に対応）が発生する原因である．

8.4 周波数領域における合焦判定法

筆者の提案手法において鍵となるのは、light field rendering 法で合成された画像上で、焦点が合っている部分をいかに検出するか、という点である．このため、筆者は第 6 章で、数種類の再構成フィルタで画像を合成し、それらの画像の差分を“合焦判定”に用いることを提案した．本章の目的は、その“合焦判定”法の枠組みの中に、ステレオマッチング法を位置づけることにある．そこで本章では、第 6 章の camera-skipped filter の考え方を拡張した、特殊なケースを考える．

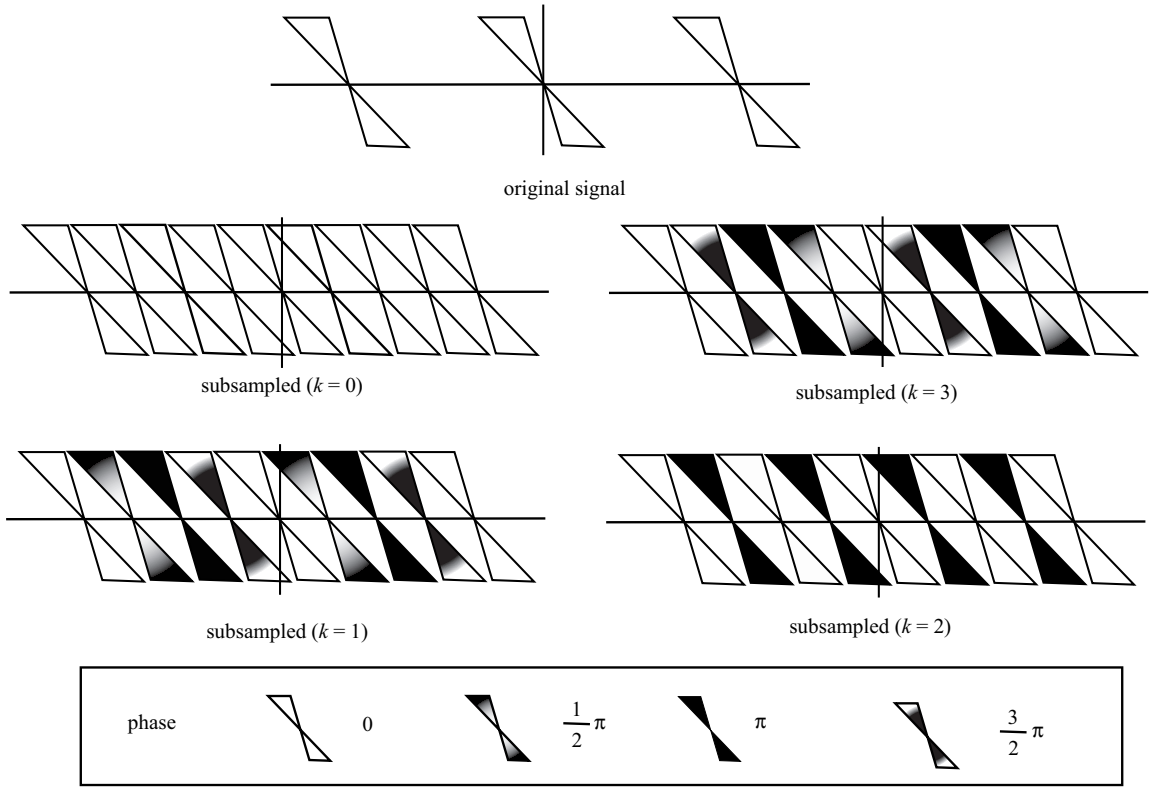
8.4.1 前提

焦点について議論したときと同様に、シーン全体ではなく、シーン中の比較的狭い領域に着目する．その領域については、奥行きの変化の幅（ $(1/z_{\min} - 1/z_{\max})$ に対応）が十分に小さく、与えられたカメラ間隔 Δs に対して、次の関係式が成り立つことを仮定する．

$$\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \leq \frac{1}{N} \cdot \frac{2\pi}{K_{\Omega_u} f \Delta s} \quad (8.3)$$

(8.1) 式よりもさらに強い条件となっているのは、“合焦判定”の過程で信号をサブサンプリングするためである． (s, u) 空間において、議論の対象となる領域を R とし、下記のような light field を定義する．

$$l(s, u) = \begin{cases} l(s, u) & (s, u) \in R \\ 0 & \text{otherwise} \end{cases} \quad (8.4)$$

図 8.3: サブサンプリングされた light field の周波数特性 ($N = 4$) .

8.4.2 light field におけるサブサンプリング

2次元の light field (s, u) 上の連続信号を $l(s, u)$, そのフーリエ変換を $L(\omega_s, \omega_u)$ とする . カメラ間隔 Δs , 画素間隔 Δu でサンプリングされた信号を $l_d(s, u)$ とし , そのフーリエ変換を $L_d(\omega_s, \omega_u)$ とする .

$$l_d(s, u) = l(s, u) \cdot \sum_{n,m} \delta(s - n\Delta s, u - m\Delta u) \quad (8.5)$$

$$L_d(\omega_s, \omega_u) = \sum_{n,m} \frac{L(\omega_s - n\Omega_s, \omega_u - m\Omega_u)}{\Delta s \Delta u} \quad (8.6)$$

$$\Omega_s = \frac{2\pi}{\Delta s}, \quad \Omega_u = \frac{2\pi}{\Delta u} \quad (8.7)$$

上記の信号を s 軸上でサブサンプリングする . サンプル間隔を N 倍として , k 番目の信号を残した信号列を $l_d^{(N,k)}(s, u)$ とする .

$$l_d^{(N,k)}(s, u) = l(s, u) \cdot \sum_{n,m} \delta(s - (nN + k)\Delta s, u - m\Delta u) \quad (8.8)$$

このフーリエ変換 $L_d^{(N,k)}(\omega_s, \omega_u)$ は次のとおりである (導出の過程は付録 1 を参照されたい.)

$$L_d^{(N,k)}(\omega_s, \omega_u) = \sum_{n,m} \left\{ \frac{L(\omega_s - n\Omega_s/N, \omega_u - m\Omega_u)}{N\Delta s \Delta u} \exp\left(-j\frac{2\pi}{N}nk\right) \right\} \quad (8.9)$$

$L_d^{(N,k)}(\omega_s, \omega_u)$ を $L_d(\omega_s, \omega_u)$ を比較すると, ω_s 軸方向の繰り返し周期が $1/N$ 倍になるとともに, $2\pi nk/N$ の位相変異が生じていることがわかる. 図 8.3 に, $N = 4$ とした場合の $L_d^{(N,k)}(\omega_s, \omega_u)$ のスペクトラムの様子を示す. 各スペクトラムの位相は図中のグラデーションで表現した凡例の通りである.

8.4.3 合焦判定の基本原理解

$L_d^{(N,k)}(\omega_s, \omega_u)$ ($k = 0, 1, \dots, N-1$) に対して, 同一の再構成フィルタ $H_{z_f}(\omega_s, \omega_u)$ を適用して, 連続信号を再構成することを考える. 例として, $L_d^{(4,0)}(\omega_s, \omega_u)$ と $L_d^{(4,1)}(\omega_s, \omega_u)$ に対して再構成フィルタを作用させる場合を, 図 8.4 に示す. ここでは, 再構成フィルタは理想低域フィルタとし, 図中に破線で示す平行四辺形の内部に帯域制限するものである.

前述したように, 再構成フィルタの傾きは, 仮定される奥行き (z_f) に対応している. 仮定される奥行きが妥当な場合 (図 8.4(a)), 両者の再構成は原理上同一の結果を与える. 一方, 仮定される奥行きが妥当ではない場合 (図 8.4(b)), 再構成の結果は異なったものとなり得る. したがって, 再構成された信号の類似度を比較すれば, その信号に対して最も妥当な奥行きを推定することができる. すなわち, 奥行き推定は, 次のような評価関数 $E(z_f)$ の最小化問題と考えることができる.

$$E(z_f) = \text{similarity} \left(L_d^{(N,k)}(\omega_s, \omega_u) H_{z_f}(\omega_s, \omega_u) \Big|_k \right) \quad (8.10)$$

具体的な評価関数としては, 差分絶対値 ((8.11) 式上段, $p = 1, 2$) や分散 ((8.11) 式下段) などが用いられる.

$$E(z_f) = \begin{cases} \sum_{k_1 \neq k_2} \left\| \iint (L_d^{(N,k_1)} - L_d^{(N,k_2)}) H_{z_f} d\omega_s d\omega_u \right\|^p & (\text{差分絶対値}) \\ \sum_k \left\{ \iint \left(L_d^{(N,k)} - \frac{1}{N} \sum_k L_d^{(N,k)} \right) H_{z_f} d\omega_s d\omega_u \right\}^2 & (\text{分散値}) \end{cases} \quad (8.11)$$

差分絶対値を用いる場合, サブサンプリングされた信号同士の差分 ($L_d^{(N,k_1)} - L_d^{(N,k_2)}$) が評価される. 分散を用いる場合, サブサンプリングされた信号と平均信号との差分 ($L_d^{(N,k)} - \frac{1}{N} \sum_k L_d^{(N,k)}$) が評価される. 付録 1 の (8.38), (8.39) 式から明らかなように, これらの差分には, サブサンプリングによって発生したエイリアシング成分のみが含まれる. したがって, $E(z_f)$ は, H_{z_f} によって制限される帯域の範囲で, それらエイリアシング成分の大きさを測る尺度とみなされる. すなわち, ここで定義する“合焦判定”法とは, 元の離散信号をサブサンプリングし, その過程で発生したエイリアシング成分を計測することで, 正しい奥行きを推定する手法と位置づけられる.

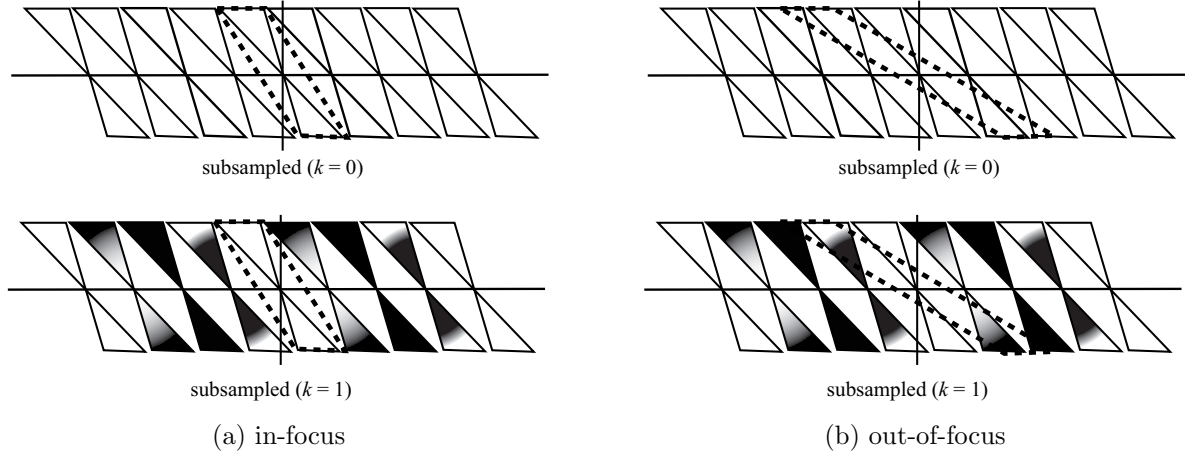


図 8.4: 再構成フィルタ：(a) 焦点が合う場合，(b) 焦点から外れる場合．

8.5 合焦判定の空間領域への展開とステレオマッチング法の導出

8.5.1 空間領域への展開

前章で導入した周波数領域における“合焦判定”の議論を，空間領域に展開する．パーシバルの等式から，“合焦”判定の評価関数 $E(z_f)$ は，空間領域では次のように表される．

$$E(z_f) = \text{similarity} \left([h_{z_f}(s, u) * l_d^{(N, k)}(s, u)] \Big|_k \right) \Big|_R \quad (8.12)$$

上記の“合焦判定”尺度は，(8.3) 式を満たすような (s, u) 空間中の領域 R に対して定義される．しかし，ある仮想視点位置で画像を合成する際には，領域 R 全体を再構成する必要はなく，その中に含まれる画像合成に必要な部分のみを再構成すれば充分である．前節で述べたように，評価関数 $E(z_f)$ は，画像合成に望ましくない効果（エイリアシング成分）の大きさを計測するものである．したがって，領域 R 全体を対象とするのではなく，所望の合成画像に寄与する部分（ R' ）に評価を限定することにより，計算コストを低減できるだけではなく，その合成画像に対して最適な評価を行うことができると考えられる．

$E(z_f)$ を計算する実装上の手順は下記の通りである．

1. 奥行き z_f を仮定して，サブサンプリングされた N 個の信号列 $l_d^{(N, k)}(s, u)$ からそれぞれ画像を合成する．それぞれの合成画像を $I_{z_f}^{(N, k)}$ と表記する．
2. 評価値 $E(z_f)$ は，次のように定義される．

$$E(z_f) = \text{similarity} \left(I_{z_f}^{(N, k)} \Big|_k \right) \Big|_{R'} \quad (8.13)$$

ここで， R' は，合成画像に寄与する R の部分空間である．

続く 2 つの節では，1. を実現する方法を説明する．画像合成は，離散的な light field をいったん補間し，そののち所望の視点位置に応じて補間されたデータをリサンプリングする処理と考えるこ

とができる．しかし，実装上は，空間領域で補間とリサンプルを同時に行うことができる．8.5.2では，再構成フィルタの空間波形を導出し，8.5.3では，その波形に基づいて画像合成を行う方法を示す．最後に8.5.4で，画像合成を零次ホールド近似で実装した場合，(8.13)式が，基本的な多眼ステレオマッチング法の評価尺度と等価になることを示す．

8.5.2 再構成フィルタの空間波形と変数分離

図8.4に示したフィルタ $H_{z_f}(\omega_s, \omega_u)$ の通過域 P は次のように表される．

$$\begin{aligned} -\frac{\pi}{N\Delta s} < \frac{f}{z_f}\omega_u + \omega_s < \frac{\pi}{N\Delta s} \\ -K_{\Omega_u} < \omega_u < K_{\Omega_u} \end{aligned} \quad (8.14)$$

理想的な低域フィルタを仮定した場合，フィルタの周波数波形 $H_{z_f}(\omega_s, \omega_u)$ は，ゲインを G として，次のように表現される．

$$H_{z_f}(\omega_s, \omega_u) = \begin{cases} G & (\omega_s, \omega_u) \in P \\ 0 & \text{otherwise} \end{cases} \quad (8.15)$$

ここで，周波数ゲイン G を下記のように定めることにより，空間波形 $h_{z_f}(s, u)$ を正規化できる．

$$G = \frac{\pi}{K_{\Omega_u}} \cdot N\Delta s \quad (8.16)$$

ところで， $H_{z_f}(\omega_s, \omega_u)$ は，二つの一次元低域フィルタ $H_1(\omega_s)$ および $H_2(\omega_u)$ の合成によって表すことができる．

$$\begin{aligned} H_{z_f} &= \left[(H_1(\omega_s)\delta(\omega_u)) * \left(H_2(\omega_u)\delta\left(\frac{f}{z_f}\omega_u + \omega_s\right) \right) \right] \\ H_1(\omega_s) &= N\Delta s \cdot H_{z_f}(\omega_s, 0) \\ H_2(\omega_u) &= \frac{\pi}{K_{\Omega_u}} \cdot H_{z_f}\left(-\frac{z_f}{f}\omega_u, \omega_u\right) \end{aligned} \quad (8.17)$$

ここで， $H_1(\omega_s)$ ， $H_2(\omega_u)$ のゲインは，それぞれの空間波形 $h_1(s)$ ， $h_2(u)$ を正規化するように定めている．上記の一次元低域フィルタを導入することにより， $H_{z_f}(\omega_s, \omega_u)$ の空間波形 $h_{z_f}(s, u)$ は，一次元フィルタ $h_1(s)$ ， $h_2(u)$ の積として表現できる（導出の詳細は付録2を参照）．

$$h_{z_f}(s, u) = h_1(s) \cdot h_2\left(u - \frac{f}{z_f}s\right) \quad (8.18)$$

ここで， u' を下記のように定義する．

$$u' = u - \frac{f}{z_f}s \quad (8.19)$$

(s, u) 空間と (s, u') 空間の変換は一意に決まる．したがって，離散信号 $l_d(s, u)$ から，連続信号 $l(s, u)$ を復元する問題は，そっくり (s, u') 空間に置き換えて考えることができる．すなわち，離

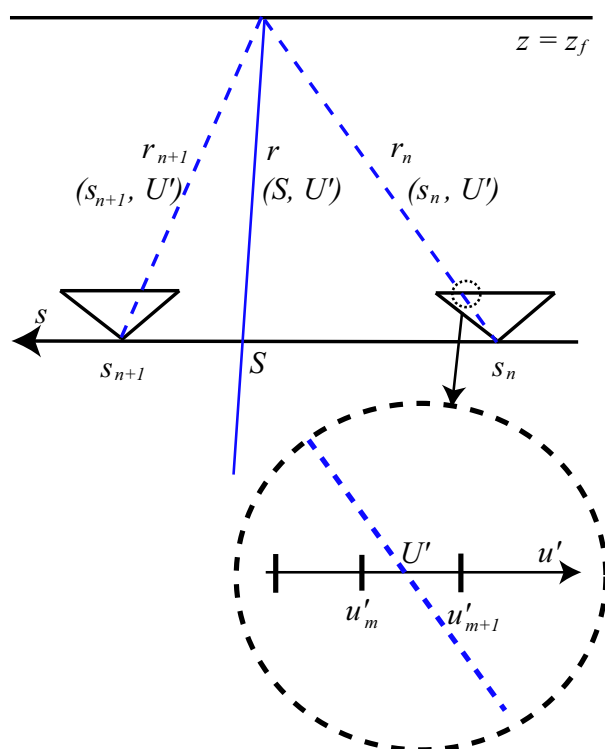


図 8.5: 画像合成の実装.

散信号 $l'_d(s, u')$ に対して，再構成フィルタ $h'(s, u') = h_1(s) \cdot h_2(u')$ を畳み込むことにより，連続信号 $l'(s, u')$ を復元することを考えればよい．すなわち，

$$\begin{aligned}
l'(s, u') &= [l'_d(s, u') * h'(s, u')] \\
&= [l'_d(s, u') * (h_1(s) \cdot h_2(u'))] \\
&= [(l'_d(s, u') * h_2(u')) * h_1(s)]
\end{aligned} \tag{8.20}$$

実装上意味を持つ重要な点は，このとき， s 軸の補間と u' 軸の補間が分離できることである．

8.5.3 画像合成の実装

本節では，上記の再構成フィルタが，オブジェクト空間においてどのように具体的に実現されるかを述べる．light field rendering では，ある仮想視点に達する個々の光線の色を求めることにより，その視点における画像を合成する．このとき，被写体の奥行きが一定であると仮定する．図 8.5 に示したように， $z = z_f$ なる奥行きを仮定し，離散的な light field データ $l_d(s, u)$ から，光線 r を合成する場合を考える（本節では，説明を簡単にするため， $N = 1$ とする）．

ここでは，光線 r の (s, u') 空間におけるパラメータを， (S, U') と表記する．(8.20) 式より， r は次の式に従って求められる．

$$\begin{aligned} r &= \left[[l'_d(s, u') * h_2(u')] * h_1(s) \right] \Big|_{s=S} \Big|_{u=U'} \\ &= \left[[l'_d(s, u') * h_2(u')] \Big|_{u'=U'} * h_1(s) \right] \Big|_{s=S} \end{aligned} \quad (8.21)$$

上記の補間のプロセスは、次の3ステップにより実現される。数学的な詳細は付録3を参照されたい。

1. 光線 r と、 $z = z_f$ なる平面の交点を求める。
2. その交点を通過して、各入力カメラに到達する光線の色を求める。 n 番目のカメラに到達する光線を、 r_n で表す。このとき、光線 r_n のパラメータ u' の値は、 U' となる。各カメラ上では、画素間の補間を行うが、これは、(8.21) 式における、 $\{l'_d(s, u') * h_2(u')|_{u'=U'}\}$ の計算に対応する。
3. 最後に、光線 r_n の重み付け和により、光線 r の色を決める。この処理は、(8.21) 式における $\{ * h_1(s)|_{s=S} \}$ の計算に対応する。

実装上、各カメラの画像をテクスチャとして扱う場合には、(2) における補間処理のために、ハードウェアのアクセラレーションを利用することが有効である（最近傍補間や線形補間は自動的に実現できる）。この場合、我々が明示的に実装するのは、(3) における補間処理（カメラ面上での補間）のみである。

8.5.4 ステレオマッチング法の導出

本節では、各画像の画面上での補間（ $h_2(u')$ による）は任意の方法で行うこととし、カメラ面上での補間（ $h_1(s)$ による）に着目する。前節の議論は、サブサンプリングされた light field データ $l_d^{(N,k)}(s, u)$ から、光線 r を再構成する場合にも容易に拡張できる。この場合には、(8.51) 式より、次式のようにカメラ面上での補間処理を行う。

$$r^{(N,k)} = \sum_{n=mN+k} h_1(S - n\Delta s) \cdot r_n \quad (8.22)$$

ここで、 m は整数である。(8.51) 式の s_n は、ここでは $n\Delta s$ と表記した。

$h_1(s)$ の波形は、 $H_1(\omega_s)$ の理想的な周波数特性をいかに近似するかに依存する。

1. 理想低域フィルタ：

$$h_1(s) = \mathcal{F}^{-1}[H_1(\omega_s)] = \frac{\sin\left(\frac{\pi}{N\Delta s}s\right)}{\frac{\pi}{N\Delta s}s} \quad (8.23)$$

2. 一次ホールド（FOH: first order hold）近似：

$$h_1(s) = \begin{cases} 1 - \frac{1}{N\Delta s}|s| & |s| \leq N\Delta s \\ 0 & \text{otherwise} \end{cases} \quad (8.24)$$

3. 零次ホールド（ZOH: zero order hold）近似：

$$h_1(s) = \begin{cases} 1 & |s| \leq \frac{N}{2}\Delta s \\ 0 & \text{otherwise} \end{cases} \quad (8.25)$$

(8.23) 式の理想低域フィルタは， $\pm\infty$ のサポートを持つため，正確に実装することは不可能である．第5，6章では，(8.24) 式の一次ホールド近似を用いた実装を採用していた．ここでは，(8.25) 式の零次ホールド近似を用いる．(8.22) 式と (8.25) 式より，光線 $r^{(N,k)}$ の色は，次のように求められる．

$$r^{(N,k)} = r_{n(k)} \quad \left(|S - n(k)\Delta s| \leq \frac{N}{2}\Delta s, \quad n(k) = mN + k \right) \quad (8.26)$$

すなわち， $n(k)$ 番目の入力カメラ画像から読み出された光線 $r_{n(k)}$ の値が，そのまま $r^{(N,k)}$ の値となる．このとき，(8.13) 式の“合焦尺度”は，下記のように表される．

$$E(z_f) = \text{similarity} \left(r_{n(k)} \Big|_k \right) \Big|_{R'} \quad (8.27)$$

この式では， N 本の光線 $r_{n(k)}$ の類似度を，領域 R' を単位として評価している．これは， $s = S$ の近傍の N 台のカメラ画像上で，奥行き z_f を仮定して「対応画素」を中心とする領域の類似度を直接評価していることに相当する．すなわち，(8.27) 式は，基本的な多眼ステレオマッチング法（オクルージョン処理や領域分割を伴わない手法）の評価値と等価である．

8.6 むすび

本章では，第5，6章で提案した，light field rendering における“合焦判定法と，ステレオ法との数学的な関係について考察した．“合焦判定法”は，多視点画像データの周波数解析理論 [2] に基づくものであり，自由視点画像合成のプロセスを，再構成フィルタによる補間としてモデル化している．本章では，元々の多視点画像データを複数の方法でサブサンプリングする場合に着目し，周波数領域での議論を，空間領域における実装へと展開した．再構成フィルタは理論上は理想低域フィルタであるが，これを零次ホールド近似によって実装した場合に，基本的な多眼ステレオ法と等価な手法が導かれることを明らかにした．

図 8.6 に示すように，“合焦”判定法は，合成画像を分析することにより，奥行き情報を得る手法である．合成の過程は，信号処理的には光線情報の補間に相当する．一般には，補間は近傍のサンプル値の重み付けによって行われるものであり，この過程で入力画像同士の混色が行われる．しかし，零次ホールドを用いた場合には，単一のサンプル値を参照する補間となるため，入力画像間の混色は行われない．したがってこれが，図 8.7 のような入力画像を直接分析するモデリング処理と等価になるというのは，原理的に妥当である．注目すべき点は，“合焦”判定が，図 8.7 を包含する枠組みを提供している，という点である．

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. Circuits and Systems for Video Technology, Vol. 13, No. 11, pp. 1020–1037 (2003).

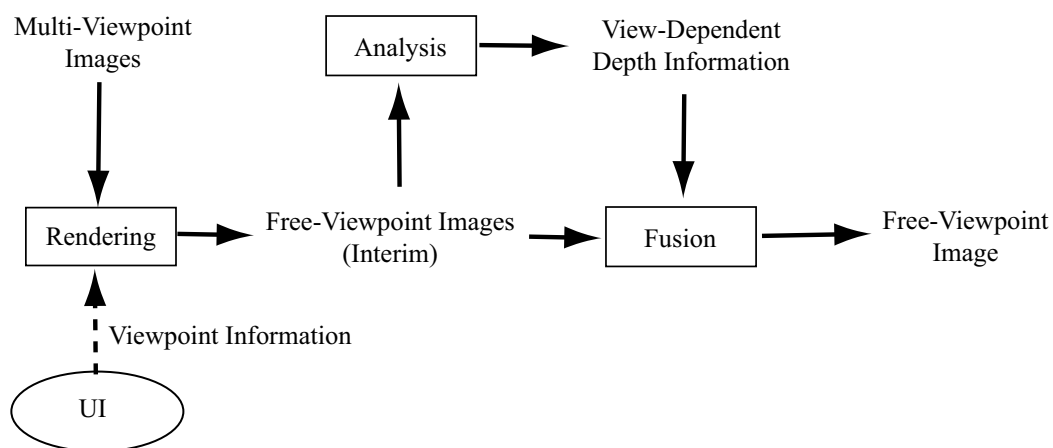


図 8.6: “合焦” 判定法の処理フロー図 .

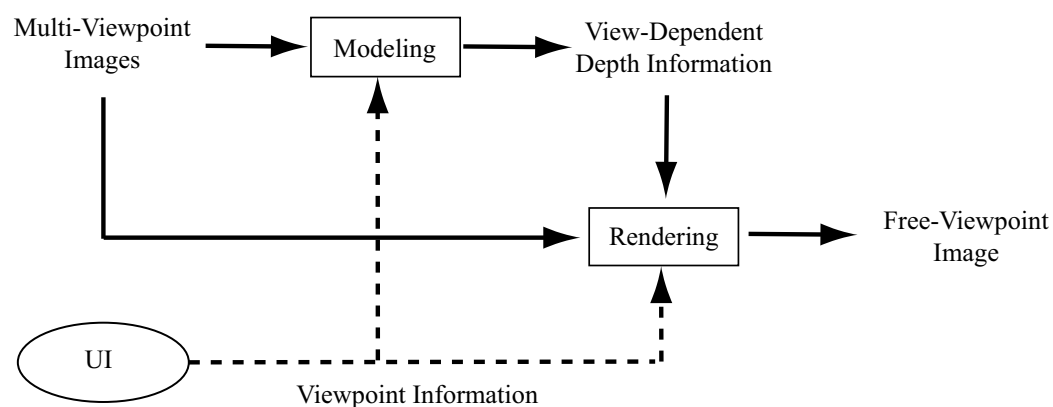


図 8.7: 視点依存型 image-based modeling and rendering の処理フロー図 .

- [2] J.-X. Chai, X. Tong, S. -C. Chany, and H. -Y. Shum, “Plenoptic Sampling”, Proc. ACM SIGGRAPH 2000, pp. 307 – 318, (2000)
- [3] U. R. Dhond and J. K. Aggarawal, “Structure from Stereo - a Review,” IEEE Trans. Systems, Man, and Cybernetics, Vol. 19, No. 6, pp. 1489–1510 (1989).
- [4] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363, 1993.
- [5] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH 96, pp. 31–42 (1996).
- [6] A. Isaksen, L. McMillan, S. J. Gortler, “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH 2000, pp. 297–306 (2000).

8.7 付録

8.7.1 サブサンプリングされた信号の周波数特性

ここでは簡単のため1次元の信号を考える．連続信号を $f(x)$ ，そのフーリエ変換を $F(\omega)$ とする．周期 X で離散化された信号 $f_d(x)$ ，およびそのフーリエ変換 $F_d(\omega)$ は次のように表される．

$$f_d(x) = f(x) \cdot \sum_{n=-\infty}^{\infty} \delta(x - nX) \quad (8.28)$$

$$F_d(\omega) = \sum_{n=-\infty}^{\infty} \frac{F(\omega - n\Omega)}{X} \quad (8.29)$$

ただし， Ω は下記によって定める．

$$\Omega = 2\pi/X \quad (8.30)$$

次に， $f_d(x)$ をサブサンプリングする場合を考える．サンプリング間隔を N 倍とし， k 番目 ($k = 0, 1, \dots, N-1$) のサンプルを残した信号系列を， $f_d^{(N,k)}(x)$ とする．

$$f_d^{(N,k)}(x) = f(x) \cdot \delta_{NX}^{(k)}(x) \quad (8.31)$$

ただし， $\delta_{NX}^{(k)}(x)$ は，周期 NX のデルタ関数列である．

$$\delta_{NX}^{(k)}(x) = \sum_{n=-\infty}^{\infty} \delta(x - X(nN + k)) \quad (8.32)$$

ここで，デルタ関数列 $\delta_{NX}^{(k)}(x)$ のフーリエ変換 $\Delta_{NX}^{(k)}(\omega)$ は，推移律によって次のように与えられる．

$$\Delta_{NX}^{(k)}(\omega) = \frac{\Omega}{N} \sum_{n=-\infty}^{\infty} \delta\left(\omega - n\frac{\Omega}{N}\right) \cdot \exp(-j\omega kX) \quad (8.33)$$

この式は，デルタ関数の部分が $\omega = n\Omega/N$ するときのみ値を持つので，次のように変形できる．

$$\begin{aligned} \Delta_{NX}^{(k)}(\omega) &= \frac{\Omega}{N} \sum_{n=-\infty}^{\infty} \delta\left(\omega - n\frac{\Omega}{N}\right) \cdot \exp\left(-j\frac{n\Omega}{N}kX\right) \\ &= \frac{\Omega}{N} \sum_{n=-\infty}^{\infty} \delta\left(\omega - n\frac{\Omega}{N}\right) \cdot \exp\left(-j\frac{2\pi}{N}nk\right) \end{aligned} \quad (8.34)$$

(8.31) 式より， $f_d^{(N,k)}(x)$ のスペクトルは， $F(\omega)$ と $\Delta_{NX}^{(k)}(\omega)$ の畳み込み積分によって与えられる．

$$F_d^{(N,k)}(\omega) = \frac{1}{2\pi} [F(\omega) * \Delta_{NX}^{(k)}(\omega)] \quad (8.35)$$

(8.34) 式を用いて，

$$F_d^{(N,k)}(\omega) = \sum_{n=-\infty}^{\infty} \frac{F(\omega - n\Omega/N)}{NX} \cdot \exp\left(-j\frac{2\pi}{N}nk\right) \quad (8.36)$$

この式は、 n 番目の複製成分には、サンプリングの位置 k に応じて、 $2\pi nk/N$ の位相変異が生じることを意味する。 n が N の整数倍になる場合は、 k の値によらず、位相は 0 である。

ここで、次のような関係 ((8.37)–(8.39) 式) が導かれる。

$$\sum_{k=0}^{N-1} F_d^{(N,k)}(\omega) = F_d(\omega) \quad (8.37)$$

この式は、 N 個のサブサンプリングされた信号列 $f_d^{(N,k)}(x)$ ($k = 0, 1, \dots, N-1$) を足し合わせると、 $f_d(x)$ が復元されることを意味する。

$$\left(F_d^{(N,k_1)}(\omega) - F_d^{(N,k_2)}(\omega) \right) \cdot F_d(\omega) = 0 \quad (k_1 \neq k_2) \quad (8.38)$$

$$\left(F_d^{(N,k)}(\omega) - \frac{1}{N} \sum_{k=0}^{N-1} F_d^{(N,k)}(\omega) \right) \cdot F_d(\omega) = 0 \quad (8.39)$$

(8.38) 式は、 $F_d^{(N,k_1)}(\omega)$ と、 $F_d^{(N,k_2)}(\omega)$ の差分には、元の離散信号 $F_d(\omega)$ と共通部分がないこと、つまり、サブサンプリングによって発生したエイリアシング成分のみが含まれることを意味する。任意の $F_d^{(N,k)}(\omega)$ と、 $F_d^{(N,k)}(\omega)$ の平均信号列についても同様である ((8.39) 式)。

8.7.2 再構成フィルタの空間波形の導出

再構成フィルタの空間波形 $h_{zf}(s, u)$ は、周波数領域波形 $H_{zf}(\omega_s, \omega_u)$ のフーリエ逆変換によって求められる。(8.17) 式と畳込み積分定理を用いることにより、

$$\begin{aligned} h_{zf}(s, u) &= \mathcal{F}_{\omega_s, \omega_u}^{-1} \left[H_{zf}(\omega_s, \omega_u) \right] \\ &= \mathcal{F}_{\omega_s, \omega_u}^{-1} \left[(H_1(\omega_s) \delta(\omega_u)) * * \left(\delta \left(\frac{f}{z_f} \omega_u + \omega_s \right) H_2(\omega_u) \right) \right] \\ &= 4\pi^2 \mathcal{F}_{\omega_s, \omega_u}^{-1} [H_1(\omega_s) \delta(\omega_u)] \cdot \mathcal{F}_{\omega_s, \omega_u}^{-1} \left[\delta \left(\frac{f}{z_f} \omega_u + \omega_s \right) H_2(\omega_u) \right] \end{aligned} \quad (8.40)$$

第一項のフーリエ逆変換は、

$$\begin{aligned} \mathcal{F}_{\omega_s, \omega_u}^{-1} [H_1(\omega_s) \delta(\omega_u)] &= \frac{1}{4\pi^2} \int_{\omega_u=0} \int H_1(\omega_s) e^{j(\omega_s s + \omega_u u)} d\omega_s d\omega_u \\ &= \frac{1}{2\pi} \int_{\omega_u=0} \left(\frac{1}{2\pi} \int H_1(\omega_s) e^{j\omega_s s} d\omega_s \right) e^{j\omega_u u} d\omega_u \\ &= \frac{1}{2\pi} \cdot h_1(s) \end{aligned} \quad (8.41)$$

第二項のフーリエ逆変換は、

$$\begin{aligned} \mathcal{F}^{-1} \left[\delta \left(\frac{f}{z_f} \omega_u + \omega_s \right) H_2(\omega_u) \right] &= \frac{1}{4\pi^2} \int \int_{\frac{f}{z_f} \omega_u + \omega_s = 0} H_2(\omega_u) e^{j(\omega_s s + \omega_u u)} d\omega_s d\omega_u \\ &= \frac{1}{2\pi} \left(\frac{1}{2\pi} \int H_2(\omega_u) \exp \left(j\omega_u \left(-\frac{f}{z_f} s + u \right) \right) d\omega_u \right) \\ &= \frac{1}{2\pi} \cdot h_2 \left(u - \frac{f}{z_f} s \right) \end{aligned} \quad (8.42)$$

(8.40)–(8.42) 式より,

$$h_{z_f}(s, u) = h_1(s) \cdot h_2\left(u - \frac{f}{z_f}s\right) \quad (8.43)$$

8.7.3 画像合成の実装についての数学的詳細

ここでは, 8.5.3 の画像合成の実装において, (2) と (3) の処理の数学的な詳細を述べる. 所望の光線 r のパラメータを, (s, u) 空間および (s, u') 空間で下記のように表記する.

$$r = l(S, U) = l'(S, U') \quad (8.44)$$

(2) について

まず, 光線 r_n のパラメータを求める. 光線 r_n は, $s = s_n$ ($s_n = n\Delta s$) を通過するので, 下記のように表される.

$$r_n = l_d(s_n, u(r_n)) \quad (8.45)$$

ここで, 幾何学的な関係から, $u(r_n)$ は次のように求められる.

$$u(r_n) = U + \frac{f}{z_f}(s_n - S) \quad (8.46)$$

上記の $(s_n, u(r_n))$ を (8.19) 式に代入する.

$$\begin{aligned} u'(r_n) &= \left(U + \frac{f}{z_f}(s_n - S)\right) - \frac{f}{z_f}s_n \\ &= U - \frac{f}{z_f}S \\ &= U' \end{aligned} \quad (8.47)$$

したがって, r_n の (s, u') 空間におけるパラメータは, 下記の通りである.

$$r_n = l'(s_n, U') \quad (8.48)$$

これは, 図 8.5 において, 光線 r_n のパラメータ u' の値が U' となることに対応する.

図 8.5 に示されるように, 離散的なサンプル集合 $l'_d(s, u')$ から (s_n, U') における光線の値を得るためには, u' 軸上での補間が必要である. したがって, 光線 r_n は, 下記の計算によって得られる.

$$r_n = [l'_d(s, u') * h_2(u')]_{u'=U'}|_{s=s_n} \quad (8.49)$$

この式から, 次の関係式が導かれる.

$$\sum_n r_n \cdot \delta(s - s_n) = [l'_d(s, u') * h_2(u')]_{u'=U'} \quad (8.50)$$

したがって, (2) において, すべての n に対して r_n を求めることにより, (8.21) 式における $\{l'_d(s, u') * h_2(u')\}_{u'=U'}$ の計算が実現されることが示された.

(3) について

(8.21) 式と (8.50) 式より光線 r は、次の式に従って求められる。

$$\begin{aligned} r &= \left[\left(\sum_n r_n \cdot \delta(s - s_n) \right) * h_1(s) \right] \Big|_{s=S} \\ &= \sum_n r_n \cdot h_1(S - s_n) \end{aligned} \quad (8.51)$$

すなわち光線 r は光線 r_n の重み付け和によって求められる。重み係数は、フィルタ h_1 の波形によって決まる。

第9章

結論

9.1 本論文の成果

近年の視覚情報メディア技術のめざましい進歩を背景に、我々が生活する三次元実環境そのものを扱う技術に期待が集まっている。本論文では、三次元視覚情報を扱う一つの有力な手段として、多眼画像を入力として自由視点画像合成を行う手法に注目して、理論と実践の両面から検討を行った。特に、既存の研究分野の枠組みにとらわれず、コンピュータビジョン、コンピュータグラフィックス、および信号処理のアプローチを融合することで、“合焦”判定法に基づく“全焦点”自由視点画像合成法を提案した。ここでは、本論文の主たる成果をまとめる。

9.1.1 多眼画像を用いた全焦点自由視点画像の実時間合成の実践

全焦点自由視点画像の実時間合成手法の提案

本論文では、キャリブレーション済みの多眼カメラによる多視点画像を入力として用いることを想定し、高品質な自由視点画像を実時間で合成する手法について検討した。第4章では、合成画像の各画素に対して、近接する四つの入力画像のみを参照して、その奥行きと色を推定することにより、実時間で自由視点画像を合成する手法を提案した。第5、6章では、light field rendering法[1]を拡張した“全焦点”自由視点画像合成法を提案した。これらの手法はいずれも、形状推定に相当する処理を内部に含み、オフライン計算を必要としないという特徴がある。

第5、6章で提案した手法の概略を述べる。light field rendering法では、被写体空間を単一の平面(focal plane)で近似するため、合成画像上では、focal planeの位置に応じて“焦点”ぼけのような現象が現れる[2]。提案手法では、focal planeの奥行きを連続的に変化させながら複数の画像を合成し、それらの画像の“合焦”領域を選択して一枚の画像に統合することで、“全焦点”自由視点画像を得る。合焦領域を選択するためには、“合焦”判定と呼ばれる独自の評価尺度を用いる。この手法をグラフィックスハードウェアのアクセラレーションを用いて効率的に実装し、実時間で画像合成が実現できることを確かめた。また、CGシーンをを用いた定量的な評価によっても、仮定するfocal planeの枚数を増やすことで、合成品質が向上することを確かめた。

全焦点自由視点画像合成法の実システムへの適用

第7章では、“合焦”判定に基づく“全焦点”自由視点画像合成手法を、実時間自由視点画像合成システムLIFLET(light field live with thousands of lenslets)[7]に適用した。LIFLETでは、レンズアレイと奥行き制御レンズという特殊な光学系を用いて、被写体空間の多眼画像が取得される。本研究では、奥行き制御レンズによる光線の屈折の影響を考慮して、画像合成手法のアルゴリズムやfocal planeの設定方法を調整した。結果として、従来用いていた「視点依存簡易モデル」による合成品質を上回る品質を達成することができた。この実験により、本研究で提案する画像合成手法が、動空間を対象とした実時間自由視点画像合成に適用可能であることが示された。

9.1.2 自由視点画像合成に関する理論的体系化

自由視点画像合成における品質限界を表す被写界深度の理論の提案

第3章では、被写体空間を表現する最も単純なモデルとして、単一の仮定平面 (focal plane) を用いた場合を想定し、被写界深度の理論を提案した。単一の仮定平面を用いる場合、合成画像上では仮定平面付近のみが鮮明となり“焦点”が合っているように見える [1, 2]。本論文では、plenoptic sampling 理論 [3] に基づき、仮定平面の前後においてエイリアシング雑音を伴わずに合成できる奥行き範囲を定量化することで被写界深度を定義した。また、この被写界深度の理論が、等価被写界深度の理論 [4] および多視点カメラシステムの解像度の理論 [5] を一般化するものであることも示した。さらに、実験により理論の定性的傾向を確認した。この検討は、最も基本的な単一の平面モデルを用いた自由視点画像合成の理論的限界を示すものであり、多眼画像システムの設計や、より複雑なモデルを用いる画像合成法の評価への応用が期待される。

全焦点自由視点画像合成のための多眼ステレオ法の最適化

第4章では、自由視点画像合成のために多眼ステレオ法 [6] を最適化する方法について検討した。動空間を対象として実時間で自由視点画像を合成するというシナリオを想定した場合、奥行き推定処理を合成画像の座標空間で直接行うことにより、計算を所望の視点位置における画像合成に必要な充分な範囲に制限できる。このアプローチのもとで、本研究では、1) シルエット抽出に依存せず、2) 画素単位の奥行き割り当てを行い、3) 領域ベースで安定した色一致度評価を行い、かつ、4) 計算量が入力カメラの台数に依存しない、最適化された奥行き推定処理を実現する手法を提案した。具体的には、合成画像の各画素ごとに、その画素を通過する光線と近接する四つのカメラのみを参照して奥行き推定を行うことで、上記の要請を満たすことができた。

合成画像の分析に基づく奥行き推定手法—合焦判定法—の提案

第5, 6章では、単一の奥行き仮定平面 (focal plane) を用いた light field rendering 法 [1] による合成画像を直接分析して“合焦”領域を検出する、“合焦”判定法を導入した。合成画像上では、focal plane の付近が“合焦”となるため、“合焦”領域を検出することは focal plane 付近の領域を検出することと本質的に同質であり、これは一種の奥行き推定である。第5章では空間領域で、6章では周波数領域で、この手法に理論的な裏づけを与えた。空間領域の議論は多眼ステレオ法と基本的原理を同じくするものであるが、任意のカメラアレイに適用でき、実用性に優れる。一方、周波数領域の議論は、理想的なカメラアレイに限定されるものであるが、多次元信号処理の理論 [3] を奥行き推定の問題に適用するユニークな試みである。

自由視点画像合成のための奥行き推定に関する理論的体系化

第8章では、多次元信号処理の観点から、“合焦”判定法と多眼ステレオ法の関係について理論的に考察した。多眼ステレオ法は入力画像群を分析する手法であるのに対し、“合焦”判定法は合成画像を分析する手法である。合成の過程は、信号処理的には光線情報の補間に相当する。一般

には、補間は近傍のサンプル値の重み付けによって行われるものであり、この過程で入力画像同士の混色が行われる。しかし、零次ホールドを用いた場合には、単一のサンプル値を参照する補間となるため、入力画像間の混色は行われない。この特殊は条件下では、“合焦”判定法は入力画像を直接分析する多眼ステレオ法と等価になる。本研究では、上記の関係を数学的に明らかにすることにより、多眼ステレオ法を包含する、自由視点画像合成のための奥行き推定に関するより広い枠組みを提唱した。

9.2 今後の課題と展望

本論文では、三次元実環境の視覚情報を表現する手段として、多眼画像を用いた自由視点画像合成技術に着目し、高品質な画像を実時間で合成する手法の提案・実践と理論的体系化を行った。本節では、本研究によって明らかになった今後の課題をまとめ、将来の展望について述べる。

9.2.1 実システムへの応用

応用に向けた詳細なチューニング

本論文では、多眼画像を入力として“全焦点”自由視点画像を合成する手法を提案・実装したが、応用に向けて、詳細なチューニングの余地がまだ多く残されている。例えば、現在の実装では、理想的な多眼画像データ（筑波大学多視点画像データベース）を用いた場合でさえ、被写体の境界部分に雑音が出やすい。これはエッジを意識せずに画像全体を一様に扱っていることに起因するものであり、今後改善の余地がある。また、理想的ではないデータを用いた場合（5章のカーネギーメロン大によるデータや7章のLIFLETによるデータ）の雑音の主な要因についても分析し、解決を図ることが必要である。

大規模カメラアレイシステムへの適用

本論文の提案手法の実システムへの適用は、現状では、比較的小規模なLIFLETシステム[7]に限定されている（第7章）。今後は、より大規模なカメラアレイを用いるシステムに対する提案手法の適用についても検討する。この目標に向けて、現在、128台のネットワークカメラを用いたカメラアレイシステム構築に関して基礎検討を始めている[8]。このカメラアレイは約80 cm × 80 cmの面積を持ち、空間を切り出す「窓」の役割を果たすものである。自由視点画像合成技術により、観察者は80 cm × 80 cmの「窓」を通して、向こうの三次元シーンを自由に眺め回すことができる。将来的には、このシステムをさらに大規模化することで、部屋の壁一面を「窓」にすることができれば、より豊かな臨場感と映像効果が実現できよう。

大規模空間を対象としたシステムへの応用

別の応用としては、コンサートホールやスタジアムなどの大規模空間を対象にした自由視点画像合成システムの構築が挙げられる。このような用途においては、大規模空間全体をカバーするようなカメラアレイを構築することは現実的ではなく、対象シーンを構造化して、モジュールご

とに扱うことが有効であると考えられる [9]。例えば、観察者からの距離に応じて遠景・中景・近景に分けたり、建物などのインフラと人物などの移動物体に分ける方法が考えられる。本論文で提案した画像合成手法は、上記のような枠組みにおいて各モジュールを扱う際には有効に適用できることが期待される。

9.2.2 理論面の発展

“合焦”判定法の最適化と多次元信号処理理論の発展

本論文の第8章では、多次元信号処理の観点から自由視点画像合成のための奥行き推定について考察し、“合焦”判定法と多眼ステレオ法を内包する枠組みを提案した。合成画像を分析する、という試みは、奥行き推定という問題に対する新たなアプローチであり、無限の可能性を持っている。しかしながら、本論文の成果は、上記の枠組みを示したことにとどまり、その枠組みの中で具体的にどんなパラメータの組み合わせが最適化であるかを示すには至らなかった。今後は、理論と実験の両面から最適化の課題に取り組む必要があると考えられる。また、本論文の信号処理的な解析（第6, 8章）は、理想的なカメラアレイ入力（各カメラに個体差がなく、厳密に二次元格子状アレイを構成する）に限定されているため、より一般的なカメラアレイを扱う信号処理理論の構築も望まれる。

物理カメラにおける焦点との理論的・方法論的な対応

別のテーマとして、本論文で検討した合成画像における“焦点”と、物理的なカメラにおける焦点について、理論的および方法論的な対応を明らかにすることが挙げられる。物理的なカメラにおいて、開口を通過する光線群は light field (in-camera light field) をなす [10]。撮像面の位置を制御することは focal plane の位置の制御に対応し、撮像は一種の light field rendering である。根本的な違いは、多眼画像から構成される light field が離散データであるのに対し、in-camera light field は連続データである点にある。児玉ら [11] は、物理カメラによる多重焦点画像群に三次元の信号処理を適用し、視点位置や焦点を制御した画像を合成する手法を報告している。これは、本論文で検討した手法と同じ目標を、異なる理論体系と方法論のもとで実現するものであり、両者の関係の解明は今後の興味深いトピックスとなろう。

以上、本研究を発展させてゆく上で明らかになった課題や、今後の可能性について述べた。三次元視覚情報処理技術の研究は、様々な研究分野にまたがって展開されており、日々目覚ましい進歩を遂げている。今後は、分野の境界を超えて技術や研究者の交流が活発になり、三次元視覚情報処理という新しい研究分野が体系化されることで、さらなる発展を遂げてゆくことが期待される。

参考文献

- [1] M. Levoy and P. Hanrahan, “Light Field Rendering,” Proc. ACM SIGGRAPH 96, pp. 31–42 (1996).
- [2] A. Isaksen, L. McMillan, S. J. Gortler, “Dynamically Reparameterized Light Fields,” Proc. ACM SIGGRAPH 2000, pp. 297–306 (2000).
- [3] J.-X. Chai, X. Tong, S. -C. Chany, and H. -Y. Shum, “Plenoptic Sampling”, Proc. ACM SIGGRAPH 2000, pp. 307 – 318, (2000)
- [4] 國田 豊, 稲見昌彦, 前田太郎, 舘 日章, “多眼カメラを用いた任意視点人物像の実時間生成システム”, 信学論, Vol. J84-D-II, No. 1, pp. 129–138 (2001)
- [5] H. Hoshino, F. Okano, and I. Yuyama, “A Study on Resolution and Aliasing for Multi-Viewpoint Image Acquisition,” IEEE Trans. on Circuits & Syst., Vol. 10, No. 3, pp. 366–375 (2000).
- [6] M. Okutomi and T. Kanade, “A Multiple-Baseline Stereo”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363, 1993.
- [7] T. Yamamoto, M. Kojima, and T. Naemura, “LIFLET: Light Field Live with Thousands of Lenslets”, ACM SIGGRAPH 2004 Emerging Technologies, etech.0130 (2004).
- [8] 高橋桂太, 田口裕一, 呉 炳俊, 一松隆平, 飯田 誠, 苗村 健, “空間共有通信のための多眼カメラアレイ構築に向けた基礎検討”, 日本バーチャルリアリティ学会第10回記念大会, pp. 475–476 (2005.9).
- [9] 通信・放送機構, “空間共有コミュニケーションプロジェクト 最終報告書”, (2002).
- [10] R. Ng, “Fourier Slice Photography,” ACM Trans. on Graphics (Proc. ACM SIGGRAPH05), pp. 735–744 (2005).
- [11] 児玉和也, 久保田彰, “3次元焦点ぼけ構造を用いて視点と焦点ぼけを操作する画像生成映像”, メディア処理シンポジウム (IMPS2005), pp. 59–60 (2005).

発表文献

研究論文（査読付き論文）

- [1] 高橋桂太, 苗村 健, 原島 博, “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. Vol. 57, No. 10, pp. 1292–1299 (2003.9).
- [2] 高橋桂太, 久保田彰, 苗村 健, “Light Field Rendering における合焦判定と全焦点画像合成”, 信学論, Vol. J88-D-II, No. 3, pp. 573–584 (2005.3).
- [3] 高橋桂太, 苗村 健, “Light Field Rendering のための合焦判定法の空間領域における定義と非格子カメラ群入力への適用”, 映情学誌, Vol. 59, No. 10, pp. 1478–1482 (2005.9).
- [4] 河 宗玄, 小島 將, 高橋桂太, 苗村 健, “レンズアレイを用いた実時間全焦点自由視点画像合成システム”, 映情学誌, Vol. 59, No. 10, pp. 1483–1487 (2005.9).
- [5] Keita Takahashi, Akira Kubota, and Takeshi Naemura, “Focus Measurement and All in-Focus Image Synthesis for Light Field Rendering”, Accepted to Systems and Computers in Japan, Vol. 37, No. 1 (2006.1) ([2] の英訳) .
- [6] Keita Takahashi and Takeshi Naemura, “Layered Light Field Rendering with Focus Measurement,” submitted to Eurasip Journals Image Communication, Special Issue on Multi-View Image Processing and Its Application in Image-Based Rendering (2006.5).

国際会議論文

- [7] Satoshi Mitsuda, Tomoyuki Yamamoto, Keita Takahashi, Takeshi Naemura, and Hiroshi Harashima, “Interactive View Synthesis from Integral Photography Using Estimated Depth Information,” Proc. SPIE Three-Dimensional TV, Video, and Display II, Vol. 5243, 5243-18, pp. 116–124, Orland, USA (2003.9).
- [8] Keita Takahashi, Takeshi Naemura, and Hiroshi Harashima, “Depth of Field in Light Field Rendering,” Proc. of IEEE Intl. Conf. on Image Processing (ICIP2003), Vol. 1, pp. 409–412, Barcelona, Spain (2003.9).

- [9] Keita Takahashi, Akira Kubota, and Takeshi Naemura, “All in-Focus View Synthesis from Under-Sampled Light Fields,” Proc. of Intl. Conf. on Artificial Reality and Telexistence (ICAT2003), pp. 249–256. Tokyo, Japan (2003.12).
- [10] Kaoru Sugita, Keita Takahashi Takeshi Naemura, and Hiroshi Harashima, “Focus Measurement on Programmable Graphics Hardware for All in-Focus Rendering from Light Fields,” Proc. of IEEE Virtual Reality 2004 Conf., pp. 255–256. Chicago, USA (2004.3)
- [11] Akira Kubota, Keita Takahashi, Kiyoharu Aizawa, and Tsuhan Chen, “All Focused Light Field Rendering,” Proc. of the Eurographics Symposium of Rendering (EGSR2004), pp. 235–242, Norrkoping, Sweden, (2004.6).
- [12] Keita Takahashi and Takeshi Naemura, “All in-Focus Light Field Viewer,” Poster presentation in ACM SIGGRAPH 2004, No. 48, Los Angeles, USA (2004.8).
- [13] Keita Takahashi, Akira Kubota, and Takeshi Naemura, “A Focus Measure for Light Field Rendering,” Proc. of IEEE Intl. Conf. on Image Processing (ICIP2004), Vol. 4, pp. 2475–2478, Singapore (2004.10).
- [14] Keita Takahashi and Takeshi Naemura, “Unstructured Light Field Rendering Using on-the-Fly Focus Measurement,” Proc. of IEEE Intl. Conf. on Multimedia & Expo. (ICME2005), pp. 205–208, Amsterdam, The Netherlands (2005.7).
- [15] Jonghyun Ha, Masaru Kojima, Keita Takahashi, and Takeshi Naemura, “All-in Focus Light Field Live with Thousands of Lenslets,” Poster presentation in ACM SIGGRAPH 2005, No. 68, Los Angeles, USA (2005.8).
- [16] Keita Takahashi and Takeshi Naemura, “Spatial Domain Analysis on the Focus Measurement for Light Field Rendering,” Proc. IEEE Intl. Conf. on Image Processing (ICIP2005), Vol. 3, pp. 13–16, Genoa, Italy (2005.9).

シンポジウム・研究会・大会等発表

- [17] 祝 捷, 高橋桂太, 牛田啓太, 苗村 健, 原島 博, “実時間 IBMR 技術を応用した電子的な鏡 (Plenoptic Digital Mirror) の試作”, 信学総大, A-16-39, p. 359 (2002.3).
- [18] Jie Zhu, Keita Takahashi, Keita Ushida, Takeshi Naemura, and Hiroshi Harashima, “Plenoptic Digital Mirror: An Electronic Mirror Using Video-Based Modeling and Rendering”, 3次元画像コンファレンス 2002, P-13, pp. 149–152 (2002.7).
- [19] 高橋桂太, 相田 仁, 齊藤忠夫, “修正サイマルキャスト法による 3次元テレビ映像符号化に向けた一検討”, 映像情報メディア学会年次大会, 10-1 (2002.8).

- [20] 高橋桂太, 杉田 馨, 苗村 健, 原島 博, “複数の距離画像を用いた実時間 IBMR システムに向けた基礎検討”, 日本バーチャルリアリティ学会第7回大会, pp. 477–480 (2002.9).
- [21] 高橋桂太, 苗村 健, 原島 博, “多眼画像取得における Plenoptic Sampling および等価被写界深度に関する理論的考察”, 画像符号化シンポジウム (PCSJ2002), P-5.01, pp. 69–70 (2002.11). ベストポスター賞および学生論文賞受賞.
- [22] 高橋桂太, 苗村 健, 原島 博, “多眼カメラアレイにおける Plenoptic Sampling と等価被写界深度に関する理論的考察”, 信学技報, IE2002-102, Vol.102, No. 469, pp. 29–34 (2002.11).
- [23] 光田智史, 山本智幸, 高橋桂太, 苗村 健, 原島 博, “奥行き推定を用いたインテグラルフォトグラフィからの自由視点画像合成”, 信学総大, D-11-150 (2003.3).
- [24] 光田智史, 山本智幸, 高橋桂太, 苗村 健, 原島 博, “実時間奥行き推定を用いたインテグラルフォトグラフィからの自由視点画像合成”, 3次元画像コンファレンス 2003, pp. 13–16 (2003.7). 優秀論文賞受賞
- [25] 高橋桂太, 苗村 健, 原島 博, “多眼画像取得におけるシーンの奥行きとエイリアシングに関する理論的考察”, 3次元画像コンファレンス 2003, pp. 133–136 (2003.7).
- [26] 高橋桂太, 久保田彰, 苗村 健, “Light Field Rendering における全焦点画像合成に向けた合焦判定に関する基礎検討”, 信学技報, PRMU2003-102, Vol. 103 No. 296, pp. 43–48 (2003.9).
- [27] 杉田 馨, 高橋桂太, 苗村 健, 原島 博, “GPU を用いた合焦判定による Light Field からの全焦点画像合成”, 信学技報, MVE2003-62, Vol. 103, No. 351, pp. 25–30 (2003.10).
- [28] 高橋桂太, 久保田彰, 苗村 健, “合焦判定を利用した Light Field からの自由視点・全焦点画像の合成”, 映像メディア処理シンポジウム 2003 (IMPS2003), pp. 15–16 (2003.11).
- [29] 櫛田和貴, 高橋桂太, 苗村 健, “視点依存簡易モデルのインタラクティブ合成による IBR の検討”, 信学総大, D-11-121 (2004.3).
- [30] 櫛田和貴, 高橋桂太, 苗村 健, “視点依存簡易モデルの実時間生成に基づく多視点画像からの自由視点画像合成”, 3次元画像コンファレンス, pp. 131–134 (2004.6).
- [31] 高橋桂太, 苗村健, “合焦判定に基づく全焦点 Light Field Viewer”, 情報科学技術フォーラム (FIT2004), 一般講演論文集 Vol. 3, pp. 221–224 (2004.9). ヤングリサーチャー賞受賞.
- [32] 高橋桂太, 苗村 健, “Light Field Rendering における合焦判定法の空間解析”, 信学技報, IE2004-101, Vol. 104, no. 493, pp. 57–62 (2004.12).
- [33] 高橋桂太, 苗村 健, “合焦判定に基づくマルチカメラ画像からの全焦点自由視点画像合成”, 計測自動制御学会 システムインテグレーション部門講演会 (SICE SI 2004), pp. 609–610 (2004.12). ベストセッション講演賞受賞.

- [34] 河 宗玄, 小島 將, 高橋桂太, 苗村 健, “レンズアレイ方式自由視点画像合成における合焦判定手法の適用”, 信学総大, D-11-143 (2005.3).
- [35] 河 宗玄, 小島 將, 高橋桂太, 苗村 健, “レンズアレイを用いた実時間全焦点自由視点画像合成システム”, 3次元画像コンファレンス 2005, pp. 17–20 (2005.7). 優秀論文賞受賞.
- [36] 高橋桂太, 苗村 健, “自由視点画像合成のための視点依存奥行きマップの実時間推定法”, 画像の認識・理解シンポジウム (MIRU2005), OS6A-24, pp. 197–204 (2005.7).
- [37] 高橋桂太, 苗村 健, “自由視点画像合成のための視点依存奥行きマップの実時間推定法”, 画像の認識・理解シンポジウム (MIRU2005), DS-19, pp. 1634–1635 (2005.7).
- [38] 苗村 健, 高橋桂太, “レンズアレイとカメラアレイを用いた全焦点自由視点画像の合成と圧縮”, 情報科学技術フォーラム 2005 (FIT2005), 3次元ビデオ処理に関するシンポジウム (2005.9).
- [39] 高橋桂太, 田口裕一, 呉 炳俊, 一松隆平, 飯田 誠, 苗村 健, “空間共有通信のための多眼カメラアレイ構築に向けた基礎検討”, 日本バーチャルリアリティ学会第10回記念大会, pp. 475–476 (2005.9).
- [40] 高橋桂太, 苗村 健, “Light Field Renderingのための合焦判定法とステレオマッチング法の関係について”, 信学技報, PRMU2005-87, Vol. 105, No. 374, pp. 19–26 (2005.10).

謝辞

修士・博士課程の5年間にわたり終始暖かいご指導をいただいた原島 博 教授に、深く感謝いたします。原島教授には、研究者として広い視野を持つことや物事の本質を見極めることの重要性を教えていただきました。また、すばらしい研究環境や、国内・海外での学会発表の場を与えていただき、思う存分研究に携わることができました。苗村 健 助教授には、研究を進める過程で、詳細な点に至るまで熱心に議論していただき、その都度、的確なご助言をいただきました。また、学会原稿執筆や学会発表のいろはを手取り足取り教えていただきました。深く感謝いたします。

本研究の基本的なアイデアは、久保田 彰 博士（東京工業大）との議論を通して生まれたものです。久保田博士とは、快く共同研究を進めさせていただきました。感謝いたします。東京大学の坂内 正夫 教授、池内 克史 教授には、博士演習の機会に貴重なご助言をいただきました。ありがとうございました。また学会の場でお世話になった下記の方々にも感謝いたします。谷本 正幸 教授（名古屋大）、齋藤 隆弘 教授（神奈川大）、相澤 清晴 教授（東京大）、Tsuhan Chen 教授（米カーネギーメロン大）、斎藤 英雄 助教授（慶応大）、亀田 能成 助教授（筑波大）、佐藤 洋一 助教授（東京大）、藤吉 弘亘 助教授（中部大）、藤井 俊彰 助教授（名古屋大）、浜本 隆之 助教授（東京理科大）、児玉 和也 助教授（国立情報学研究所）、國田 豊 博士（(株)NTT）、神原 誠之 博士（奈良先端大）、大井 隆太郎 博士（NHK）、山崎 俊彦 博士（東京大）。

原島・苗村研究室で、筆者と共に三次元画像技術の研究に携わり、議論や苦労を共にした下記の方々に感謝いたします。杉田 馨 博士（現在、(株)東芝）、祝 捷 氏（現在、米コーネル大博士課程）、山本 智幸 氏（現在、(株)シャープ）、光田 智史 氏（現在、(株)NTT データ）、小島 將 氏（現在、NHK）、櫛田 和貴 氏（現在、東京大学石塚研究室）、田口 裕一 氏（修士2年）、呉 炳俊 氏（現在、米コロンビア大修士課程）、河 宗玄 氏（修士1年）、一松 隆平 氏（学部4年）、金 時煥 氏（学部4年）。また、飯田 誠 博士（助手）、田中 崇 氏（助手）、および事務補佐員の方々には、研究生活においてお世話になりました。柿本 正憲 博士（現在、(株)日本 SGI）、ナイワラ P. チャンドラシリ 博士（研究員）、牛田 啓太 博士（現在、群馬高専）、森村 久美子 氏（博士3年）、加藤 寛 氏（現在、(株)NTT ドコモ）、Dominick Chen 氏（修士2年）には、学会原稿の校正をしていただきました。ありがとうございました。

すべての方のお名前を挙げることはできませんでしたが、同輩や、研究室で生活を共にした方々、そのほか筆者を支えてくださった方々に、この場を借りてお礼申し上げ、本論文の締めくくりとさせていただきます。

2005 年 12 月 16 日