

修士論文

自由視点画像合成に基づく
3次元空間情報の視点依存型符号化の研究

2006年2月3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 46417

田口 裕一

内容梗概

実世界の3次元空間の視覚情報を扱う技術の1つとして、Image-Based Rendering (IBR) の研究が近年盛んに行われている。IBR では、ある空間の多視点画像を入力として、その空間を自由な視点から見た画像(自由視点画像)を合成する。臨場感あふれる3次元空間情報の記録・再生を目的としたこの技術は、現在の2次元情報による映像表現を大きく飛躍させるための技術として注目されている。

IBR における1つの技術的な課題として、膨大な情報量を持つ入力多視点画像群をいかに効率的に符号化するかという問題がある。現在の2次元画像システムにおいても、ほとんどの場面において符号化技術が利用されていることからわかるように、符号化技術は画像データの伝送・蓄積の際には必要不可欠なものである。多視点画像の符号化においては、1視点のみの情報である2次元画像よりも、さらに効率的な符号化技術が求められる。また、IBR において、入力多視点画像データは最終的に画像合成に利用されるため、圧縮効率は勿論のこと、画像合成時のコストやアプリケーションの利用形態などを総合的に考慮した、機能的な符号化が望まれる。

本研究では、特に平面アレイ状に撮影される入力多視点画像を利用して自由視点画像合成を行うシステムを想定する。このようなシステムにおいては、最終的に提示される画像はある視点位置での2次元合成画像であるため、低ビットレートにおいては、視点位置に対して優先度を設定し、その視点付近の情報を重点的に符号化することが有効である。一方、高いビットレートでデータを伝送できる場合、より多数の視点位置における画像を高品質に合成できるような、スケーラビリティを持つ符号化が有用であると考えられる。

そこで本研究では、自由視点画像合成技術を符号化の前処理として利用することで、任意の視点位置での2次元合成画像を中心とした階層性を実現する、スケーラブルな符号化手法を提案する。提案手法では、まず視域の中から1つの視点位置を自由に選び、それを「代表視点位置」と定める。そして、その視点における2次元画像をキーとして符号化を行う。これは、代表視点位置およびその周辺の情報を重点的に符号化し、それ以外の視点位置における情報をビットレートに応じて付加していく、階層的な符号化方式である。すなわち、復元可能な視点数と各視点位置における合成画像品質が、ビットレートの向上に伴って改善されるようなスケーラビリティを実現する。

評価実験においては、2つの異なる多眼撮像系により取得されたデータ(カメラアレイ入力およびインテグラルフォトグラフィ入力)を利用して提案手法の有効性を検証する。これにより、提案手法では高い圧縮効率と視点依存階層性という機能性の両立が可能であることを示す。

目次

第 1 章	序論	1
1.1	本研究の背景と目的	2
1.2	本論文の構成	2
第 2 章	関連研究	4
2.1	はじめに	5
2.2	光線情報による空間記述	5
2.2.1	光線空間	5
2.2.2	Image-Based Rendering のための光線空間の表現	5
2.3	Image-Based Rendering	6
2.3.1	Light Field Rendering	6
2.3.2	幾何モデルの導入による合成画像品質の改善	7
2.3.3	平面アレイ型多眼撮像系システム	7
2.4	光線空間の符号化	11
2.4.1	光線空間における相関	11
2.4.2	光線空間符号化に対する要請	11
2.5	光線空間符号化の具体的手法	12
2.5.1	画像群を多次元光線空間データとして均等に扱う手法	12
2.5.2	画像間の予測をブロック探索ベースで行う手法	13
2.5.3	画像間の予測を幾何モデルベースで行う手法	15
2.5.4	符号化手法のまとめ	18
2.6	まとめ	19
第 3 章	自由視点画像合成に基づく光線空間符号化手法の提案	20
3.1	はじめに	21
3.2	提案符号化手法の概要と目的	21
3.3	符号化手順	22
3.4	階層的符号ストリーム	23
3.5	提案手法の評価方法	23
3.6	関連研究との比較	24
3.7	まとめ	25
第 4 章	カメラアレイ入力およびインテグラルフォトグラフィ入力に対する実装	26
4.1	はじめに	27
4.2	符号化器の実装	27
4.2.1	画像合成と視点依存幾何モデルの構築	27
4.2.2	予測画像の合成と符号化モードの選択	29

4.2.3	階層的ビットストリームの構成	29
4.3	提案符号ストリームを用いたアプリケーション	30
4.3.1	JPEG 互換性を持つ符号ストリーム	30
4.3.2	代表視点画像と視点依存幾何モデルを利用したレンダリング	30
4.4	まとめ	33
第 5 章	符号化効率と合成画像品質の視点依存性の評価	34
5.1	はじめに	35
5.2	実験条件	35
5.3	代表視点位置による予測効率の変化	38
5.4	入力画像群のレート歪特性	41
5.5	合成画像品質の視点依存性	44
5.5.1	評価手法の詳細	44
5.5.2	実験結果	45
5.5.3	復元される合成画像の視覚特性	47
5.6	レンダリング手法による合成画像品質の違い	47
5.7	まとめ	50
第 6 章	結論	51
6.1	本研究の成果	52
6.2	今後の展望	52
	謝辞	55
	参考文献	56
	発表文献	61

目次

2.1	仮想視点位置における画像合成	6
2.2	Light field rendering	7
2.3	エイリアシングを生じないための画像取得密度とレイヤの枚数の関係	8
2.4	カメラアレイの例	9
2.5	LIFLET 撮像系	10
2.6	LIFLET により撮影されるインテグラルフォトグラフィと合成画像の例	10
2.7	階層的な多視点画像の符号化順序	14
2.8	視差補償予測と動き補償予測を組み合わせた多視点動画の符号化	14
2.9	多視点画像とエビポーラ平面画像	16
2.10	Model-Aided Coding における予測	17
2.11	Surface light field	18
3.1	提案符号化手法が目的とするアプリケーション例	21
3.2	提案符号化手法のブロック図	22
3.3	階層的符号ストリーム構造	23
3.4	合成画像品質の視点依存性の評価方法	24
4.1	実装の詳細とインテグラルフォトグラフィを利用した画像例	27
4.2	カメラアレイ入力に対する画像合成法	28
4.3	インテグラルフォトグラフィ入力に対する画像合成法	29
4.4	JPEG 互換性を持つ符号ストリームの利用	31
4.5	カメラアレイ多視点画像を利用したレンダリング例	32
5.1	カメラアレイ入力多視点画像の一部	36
5.2	インテグラルフォトグラフィ入力画像	37
5.3	インテグラルフォトグラフィの要素画像に対するパディング処理	38
5.4	座標系の設定	39
5.5	視点位置による合成画像の違い	39
5.6	代表視点位置を前後に動かした場合の予測画像品質の変化	40
5.8	MPEG-2 符号化器におけるスキャン順序	41
5.7	レート歪特性を求める際の代表視点画像	42
5.9	入力画像群のレート歪特性	43
5.10	合成画像品質の視点依存性の評価方法	44
5.11	合成画像品質の視点依存性	46
5.12	原画像の品質が低い視点位置の例	47
5.13	復元される合成画像例	48
5.14	各種レンダリング方法を用いた場合の合成画像品質	49

6.1	画像合成に利用される入力画像中の参照領域	53
6.2	カメラアレイプロトタイプの外観	54

表目次

2.1	光線空間符号化手法のまとめ	19
4.1	JPEG 拡張領域に格納されるデータ	31
5.1	入力多視点画像群のパラメータ	35
5.2	予測効率を最大化する代表視点位置	42

第1章

序論

1.1 本研究の背景と目的

実世界の視覚情報を伝送することは、極めて直感的で有力なコミュニケーション手段である。写真やビデオなど、視覚情報を記録・再生するための技術がこれまでに考案され、広く普及してきた。また、通信技術の進歩により、インターネットを介したテレビ会議なども可能となった。しかし、現在の技術は、本来3次元空間である実世界を、2次元の静止画や動画として切り出したものに過ぎない。より臨場感のある視覚情報通信を実現するためには、3次元空間情報をありのままに記録・再生する技術が望まれる。

3次元空間の視覚情報を扱う技術の1つとして、Image-Based Rendering (IBR) の研究が近年盛んに行われている。IBRでは、ある空間の多視点画像を入力として、その空間を自由な視点から見た画像(自由視点画像)を合成する。実物さながらのオブジェクトを様々な視点から眺めることや、3次元空間内を自由に歩き回るなど、臨場感あふれる3次元空間情報の提示を可能とするこの技術は、現在の2次元情報による映像表現を大きく飛躍させるための技術として、映画、エンターテインメント、遠隔地コミュニケーションなど、様々な分野への応用が期待されている。

IBRにおける1つの技術的な課題として、膨大な情報量を持つ入力多視点画像群をいかに効率的に扱うかという問題がある。現在の2次元画像システムにおいても、ほとんどの場面において符号化技術が利用されていることからわかるように、符号化技術は画像データの伝送・蓄積の際には必要不可欠なものである。多視点画像の符号化においては、1視点のみの情報である2次元画像よりも、さらに効率的な符号化技術が求められる。また、IBRにおいて、入力多視点画像データは最終的に画像合成に利用されるため、提示段階において扱いづらい符号化形式は望まれない。このため、圧縮効率は勿論のこと、画像合成時のコストやアプリケーションの利用形態などを総合的に考慮した、機能的な符号化が必要となる。

本研究では、特に平面アレイ状に撮影される入力画像群を利用して自由視点画像合成を行うシステムを想定する。このようなシステムにおいては、最終的に提示される画像はある視点位置での2次元合成画像であるため、低ビットレートにおいては、視点位置に対して優先度を設定し、その視点付近の情報を重点的に符号化することが有効である。一方、高いビットレートでデータを伝送できる場合、より多数の視点位置における画像を高品質に合成できるような、スケーラビリティを持つ符号化が有用であると考えられる。

そこで本研究では、自由視点画像合成技術を符号化の前処理として利用することで、任意の視点位置での2次元合成画像を中心とした階層性を実現する、スケーラブルな符号化手法を提案する。提案手法では、まず視域の中から1つの視点位置を自由に選び、それを「代表視点位置」と定める。そして、その視点における2次元画像をキーとして符号化を行う。これは、代表視点位置およびその周辺の情報を重点的に符号化し、それ以外の視点位置における情報をビットレートに応じて付加していく、階層的な符号化方式である。すなわち、復元可能な視点数と各視点位置における合成画像品質が、ビットレートの向上に伴って改善されるようなスケーラビリティを実現する。

1.2 本論文の構成

本論文の構成は以下の通りである。

第2章では、関連研究を紹介する。IBR手法およびその符号化方式を体系的に説明するために、まず「光線情報による空間記述」という概念を導入する。IBRにおける多視点画像データを、光線空間データとして考えることにより、より一般的な枠組みの中で議論を展開できる。次に、IBRの基礎的な画像合成手法と入力システムとなる多眼撮像系について説明する。最後に、既存の光線空間符号化手法を、基本的な考え方や機能性の観点から体系的にまとめる。

第3章では、自由視点画像合成に基づく光線空間データの視点依存型符号化手法を提案する。これは、ある視点位置およびその周辺の情報を重点的に符号化し、それ以外の情報をビットレートに応じて付加し

ていく，階層的な符号化方式である．ここでは，提案手法が想定するアプリケーション例について説明し，提案手法における符号化手順の概要について述べる．また，このような手法に適した評価方法について検討するとともに，関連研究と比較した場合の提案手法の位置づけを述べる．

第4章では，カメラアレイ入力およびインテグラルフォトグラフィ入力に対する，符号化器の実装の詳細について述べる．また，これにより出力される階層的な符号ストリームの利用方法について，具体的なアプリケーションを交えて説明する．

第5章では，提案手法の評価実験を行う．まず，代表視点位置に応じて予測効率が変化することを示し，符号化効率を最大化する観点での代表視点位置の選び方について検討する．次に，提案手法のレート歪特性を，一般的な2つの符号化手法と比較しながら評価する．提案手法では，視点依存階層性という機能性に加えて，高い圧縮効率も達成されることが示される．最後に，合成画像品質の視点依存性を評価し，提案手法の機能性を検証する．

第6章では，本研究の成果をまとめるとともに，今後の展望について述べる．

第2章

関連研究

2.1 はじめに

Image-Based Rendering (IBR) とは、ある空間を多数の視点から撮影した実写画像群をもとに、その空間を自由な視点位置から見た、自由視点画像を合成する技術である。IBR には数多くの手法が存在するが [1, 2], 「光線情報による空間記述 [3-5]」という概念を導入することにより、これらを体系的に説明することができる。「光線情報による空間記述」では、3次元空間を光線の集合(光線空間)として表現する。視覚情報は目に入射する光線の情報によって得られるので、任意視点での画像を合成することは、その視点に入射する光線の情報を再現することに等しい。つまり IBR は、画像の撮影によってサンプリングされた実空間の光線情報をもとに光線空間を構築し、その空間から光線情報を読み出すことで画像の合成を行う手法と考えることができる。

このように、実写画像から直接光線を合成する IBR 技術を利用すると、オブジェクトの幾何モデル、反射特性、照明の位置などから擬似的に光線を生成する伝統的なコンピュータグラフィックスに比べて、コストの低い計算でより写実性の高い画像を合成することが可能となる。しかし、IBR における光線空間の実体は膨大な数の静止画像群や動画群であり、取得したデータをそのまま扱うことは極めて困難である。

このため、効率的な圧縮符号化手法が必要不可欠であり、数多くの研究が行われている。光線空間の符号化では、従来からの手法である静止画、動画の符号化手法ももちろん利用されるが、被写体の幾何モデルを導入することにより、さらに効率の良い符号化が実現されている。また、符号化方式にランダムアクセス性やプログレッシブ性などの性質を持たせることにより、インタラクティブレートでの画像合成を行う研究も盛んである。

本章では、IBR の研究動向に関して、特に光線空間の符号化方式に着目して議論を行う。2.2 節では光線による空間記述と IBR の関係について述べる。2.3 節では IBR の概要と、最近の多眼撮像系システムを紹介する。2.4 節で光線空間における相関と符号化に必要なとされる性質について述べ、2.5 節でこれまでに提案されてきた具体的符号化手法について説明する。最後に 2.6 節でまとめを行う。

2.2 光線情報による空間記述

2.2.1 光線空間

視覚情報は空間を伝搬する光によって得られるので、3次元空間を満たす光線の情報を全て伝送することができれば、3次元空間の視覚情報の伝送が達成されることになる。このような3次元空間内の全光線情報は、任意の位置 (x, y, z) における (θ, ϕ) 方向からの光線情報(輝度や色)によって記述できる。つまり、

$$f(x, y, z, \theta, \phi) \quad (2.1)$$

のように5次元の関数として表現される [5]。対象となる空間が時間的に変化する動空間の場合は、これに時間変数 t を加えた6次元の関数を考えればよい。このような関数によって記述される空間を光線空間と呼ぶ。

2.2.2 Image-Based Rendering のための光線空間の表現

5次元もしくは6次元の光線情報は、そのままでは膨大なデータ量となり扱いづらい。また、IBR における画像取得の段階において、3次元実空間中の全ての点における光線情報を集めることは実用上不可能である。

このため、多くの IBR の手法においては、図 2.1 のように、3次元空間を被写体がある空間と観察者の視点がある空間に分割して考える。「光が3次元空間内を干渉や減衰などの影響を受けずに直進する」と仮

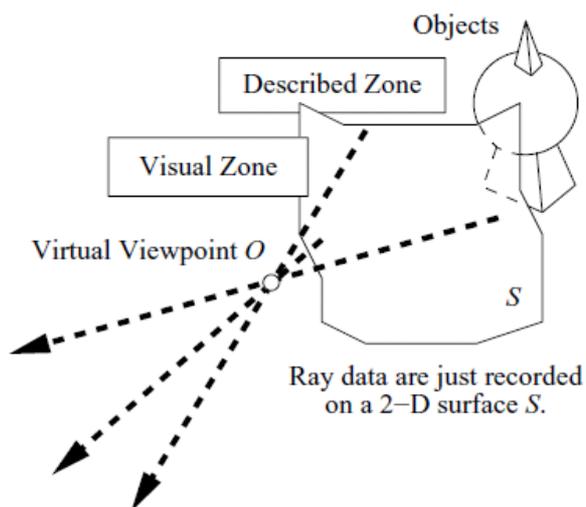


図 2.1: 仮想視点位置における画像合成 (文献 [6] より)

定すると、境界面 S 上の光線情報から任意視点位置 O での画像を合成できる。つまり、境界面 S 上の光線情報のみを取得すればよく、この場合光線空間は

$$f(p, q, \theta, \phi) \mid (x, y, z) = S(p, q) \quad (2.2)$$

のように 4 次元の関数として表すことができる [6]。

このような記述方式では、視点位置の自由度は 3 次元となる。また、 S が被写体を囲むように配置された場合には、物体を外から眺め回すための物体記述、 S が観察者の視点を囲むように配置された場合には、領域内から外の環境を眺めるための環境記述を行うことができる。つまり、(2.2) 式のように 4 つのパラメータがあれば、十分に自由度の高い空間記述を実現できる。また、(2.1) 式と比べて次元を 1 つ減らすことができるという意味において、効率的な空間記述になっていることがわかる。

2.3 Image-Based Rendering

本節では、IBR における画像の取得、および任意視点画像合成の概要を説明する。ここでは、特に平面アレイ状に多視点画像を取得するシステムを対象とする。まず、そのようなシステムにおける画像合成法の最も基本となる Light Field Rendering [7] について述べる。また、幾何モデルの導入による合成画像品質の改善について説明し、最後に現在開発されている多眼撮像システムについて紹介する。取得した画像の符号化に関しては、2.4 節以降で詳しく説明する。

2.3.1 Light Field Rendering

Levoy ら [7] によって提案された Light Field Rendering (LFR) は、IBR の基本的な手法の 1 つであり、以後の多くの研究に影響を与えている。

LFR では、2 次元格子状に並べられたカメラを利用して画像を取得する。つまり、図 2.1 の境界面 S として 2 次元平面を利用していることに相当する。

画像合成は、仮定された視点位置に到達する全ての光線情報を求めることによって行われるが、実際に撮影されていない光線については補間処理を行う必要がある。この際、図 2.2 のように被写体空間を 1 枚

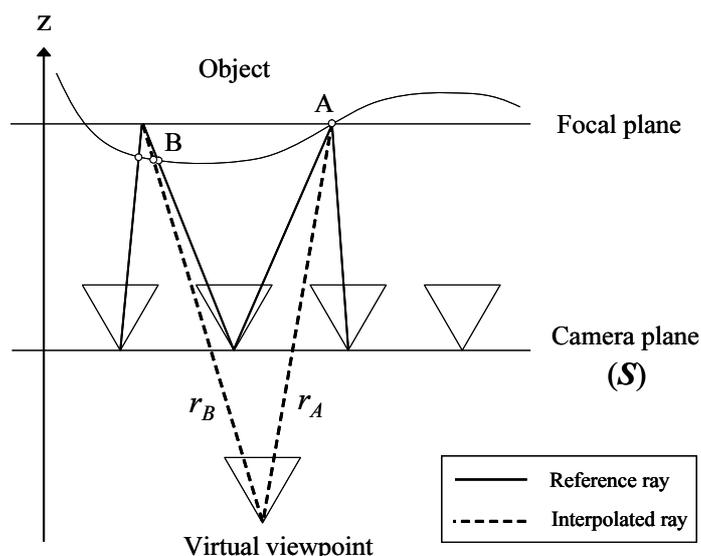


図 2.2: Light field rendering

の平面 (focal plane) で近似し、全ての光線がその平面上から発せられていると仮定する。例えば、A 地点から発せられる光線 r_A を求める場合、 r_A と camera plane の交点の近傍にあるカメラで取得された光線を参照して補間が行われる。

しかし、focal plane 上にはない B 地点からの光線 r_B を同様に求めようとした場合、オブジェクト上の別の地点からの光線を参照してしまうため、正しい補間を行うことができない。このように、LFR では focal plane 付近のオブジェクトは正しく再現できるが、focal plane から離れたオブジェクトにはぼけや 2 重像が発生することが知られている [8]。

2.3.2 幾何モデルの導入による合成画像品質の改善

2.3.1 節で述べたぼけや 2 重像の問題を回避するためには、カメラを十分に密に配置し光線空間の密なサンプリングを行うか、幾何モデルを導入して精度の良い補間を行う必要がある。光線空間の十分なサンプリングは、データ量やカメラの物理的な大きさによる制約などから実用上不可能であり、通常は何らかの幾何モデルが導入されることが多い [9]。

Chai ら [10] は幾何モデルを奥行き方向の多層平面レイヤで近似した場合、エイリアシングを生じない画像合成を行うために必要な画像のサンプル数とレイヤの数が反比例になることを理論的に示している (図 2.3)。これは、システム設計において光線情報のサンプリング密度と奥行き情報の詳細度とのトレードオフを考える際の指針となるものである。

2.3.3 平面アレイ型多眼撮像系システム

初期の IBR の研究は主に静止画を対象としていたが、計算機や撮像機器の高性能化に伴い、光線空間の取得から自由視点画像の提示までを実時間で行うことを目的としたシステムが多数提案されている。ここでは、カメラアレイおよびレンズアレイを利用した多眼撮像系システムについて紹介する。

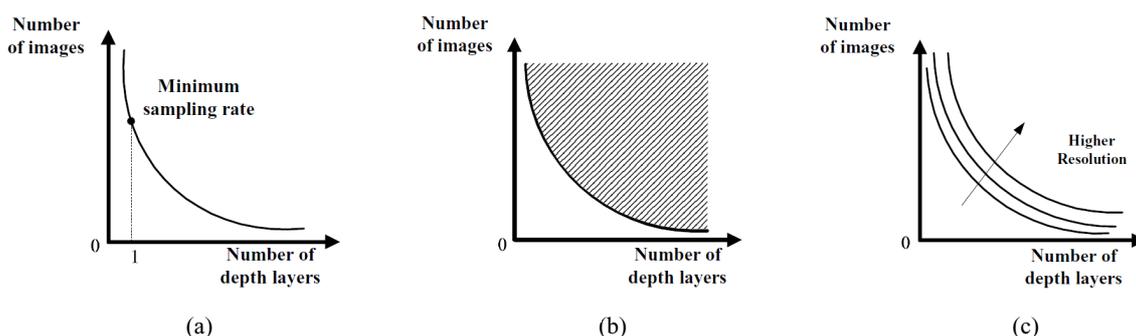


図 2.3: エイリアシングを生じないための画像取得密度とレイヤの枚数の関係 (文献 [10] より)

カメラアレイ撮像系

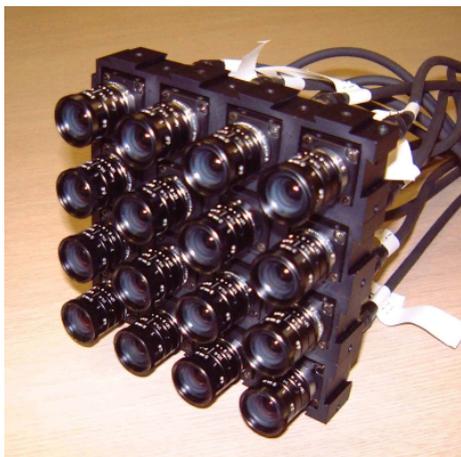
図 2.4 は、自由視点画像合成システムの撮像系として用いられているカメラアレイをまとめて示したものである。初期のカメラアレイは、図 2.4 (a) に示すような、NTSC カメラを用いた小規模なものであった [11]。近年では、カメラの低廉化・小型化および計算機の高性能化に伴い、撮影可能な被写体空間の拡大と合成画像の品質の向上を目標として、大規模なカメラアレイシステムが開発されている。

現在のところ、画素数とカメラ数の積を考えると、最も大規模なものは Tanimoto [16, 17] らのカメラアレイである (図 2.4 (e))。カメラの台数は 100 台、各カメラの解像度は 1394×1040 画素である。一方、カメラ数のみでは、Wilburn ら [14, 15] が 128 台のカメラアレイを構築している (図 2.4 (d))。Wilburn らはこのカメラアレイを用いて、自由視点画像合成だけでなく、高解像度、高フレームレート、特殊開口撮影などの検討も行っている [15]。ただし、これら 2 つのシステムでは、取得から合成までをリアルタイムに行うという報告はなされていない。

Zhang ら [12] は、カメラの台数を増やすのではなく、それぞれのカメラを動的に再配置することにより、合成画像品質を向上させるシステムを提案している。48 台のカメラを利用したこのシステム (図 2.4 (b)) では、各々のカメラにはサーボモータが付けられており、水平方向への移動とパン (左右への首振り) 動作を行うことができる。カメラの位置は固定ではないので、画像合成の際には、まずキャリブレーションによりカメラの位置と向きを計算する必要がある。次に、被写体の幾何モデルを推定するために、color-consistency [18, 19] の評価を、合成画像上においてある解像度を持つメッシュの頂点上で行う。あるメッシュの頂点において、推定された奥行きが極度に違う場合、そのメッシュを階層的に分割して奥行きの推定を続ける。最後に、Buehler ら [20] のアルゴリズムを用いて合成画像のレンダリングが行われる。この過程において、被写体の幾何モデルには信頼度が高い部分と低い部分が存在する。例えばオブジェクトと背景との境界など、オクルージョンにより構造推定が難しい部分は信頼度が低い。この部分の信頼度を向上させるために、カメラを制御し自動的にこの部分に集中させるアルゴリズムが用いられる。これにより幾何モデルの精度を向上させることができ、高品質の合成画像を提示することができる。

レンズアレイ撮像系

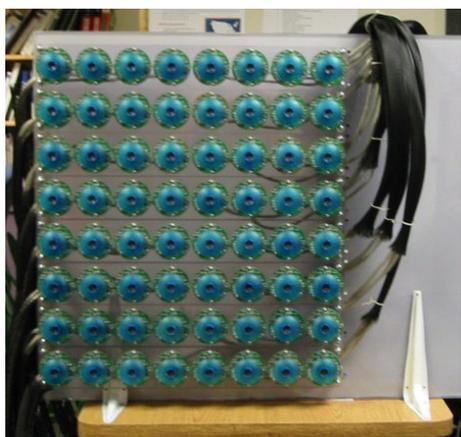
Yamamoto ら [21] は、カメラアレイの代わりにレンズアレイと 1 台のカメラを撮像系に利用したシステム (LIFLET) を構築している (図 2.5)。レンズアレイを利用することは、小さなカメラを密に並べることと等価であるが、カメラアレイを用いる場合に比べ、より多数の視点から、より密な光線情報を取得できるという利点がある。LIFLET により撮影される画像を図 2.6 (a) に示す。この画像はインテグラルフォトグラフィ [22] と呼ばれ、1 つのレンズに対応する小さな円領域の画像は要素画像と呼ばれる。画像合成の際に



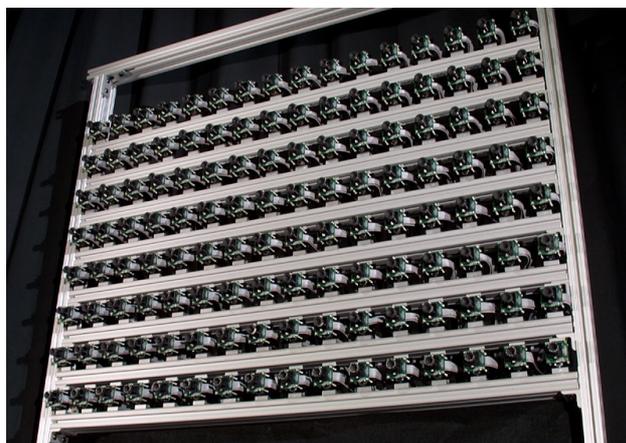
(a) 16 cameras [11]



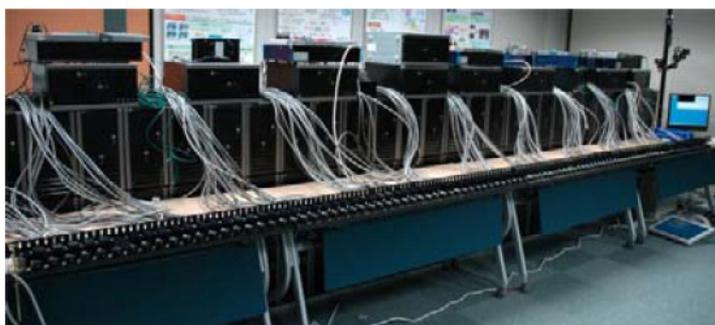
(b) 48 cameras [12]



(c) 64 cameras [13]



(d) 128 cameras [14, 15]



(e) 100 cameras [16, 17]



図 2.4: カメラアレイの例

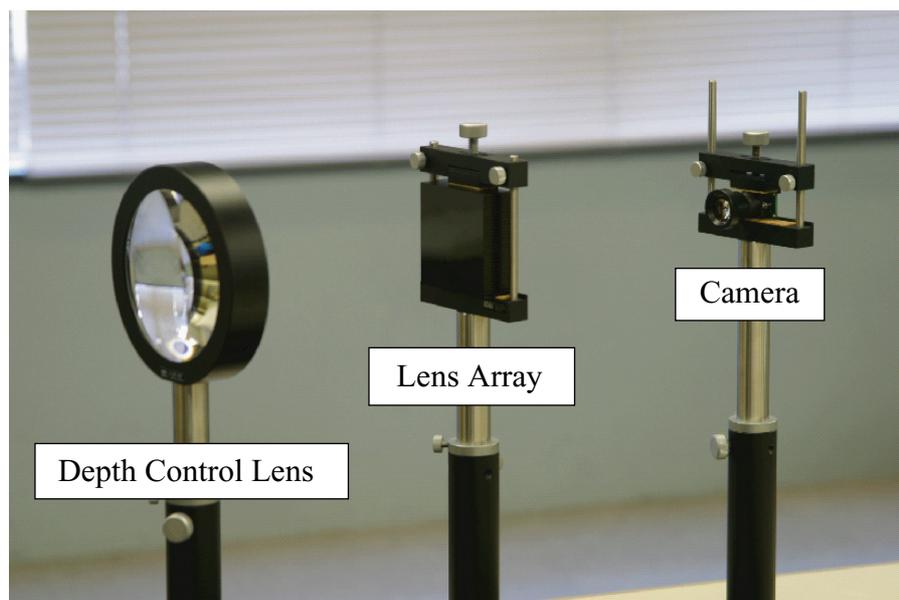
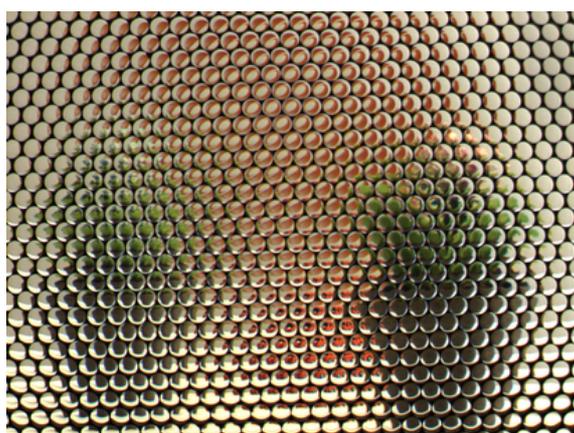


図 2.5: LIFLET 撮像系



(a) An integral photography



(b) A synthesized image

図 2.6: LIFLET により撮影されるインテグラルフォトグラフィと合成画像の例

は、まず、ステレオマッチングを利用して各レンズごとに1つの奥行きを推定することにより、視点依存簡易幾何モデルを構築する。これを用いることにより、図 2.6 (b) のような高品質な画像をインタラクティブプレートで提示することに成功している。

小島らは、この撮像系の空間周波数特性を解析し [23]、カメラの数を増やすことでより高品質な画像を合成するシステムを提案している [24]。また、河ら [25] は、カメラアレイ入力に対して提案されていた合焦判定法 [26,27] をこのシステムに適用することにより、ステレオマッチングを用いた手法よりも高品質な画像を合成できることを示した。

2.4 光線空間の符号化

IBRにおける光線空間の実体は膨大な情報量の静止画像群や動画画像群であるので、実用的な利用においては情報を圧縮することが必要不可欠である。幸い光線空間には多くの相関が含まれているので、これらを有効に利用することにより、効率的な圧縮が実現できる。本節では、光線空間に含まれる相関と、光線空間符号化に求められる性質について述べる。

2.4.1 光線空間における相関

光線空間には、以下のような4つの相関が含まれている。

- (1) 光線の伝播経路における相関
- (2) 隣接光線情報の相関
- (3) 時系列光線情報の相関
- (4) 空間中の1点に対応する光線情報の相関

まず(1)の相関については、2.2.2節における「光が3次元空間内を干渉や減衰などの影響を受けずに直進する」という仮定において利用されている。これにより、光線空間の効率的な4次元表現が可能になることはすでに示したとおりである。

(2)および(3)の相関は、従来の静止画像および動画の圧縮にそれぞれ利用されているものである。つまり、従来画像処理の分野で培われてきたJPEGやMPEGなどの手法を、光線空間の圧縮に直接適用することもできる。ただしそれだけでは、光線空間の冗長性を最大限に利用することはできない。

(4)の相関は光線空間独特のものであり、オブジェクト上の1点はどの方向から見ても同じ色に見える場合が多いことを意味している。後ほど説明する、隣接視点間でブロック探索を利用して予測を行う手法は、この相関を暗に利用しているといえる。また、より明示的にこの相関を利用する手法としては、オブジェクトの幾何モデルを構築する方法がある。これらの手法は、光線空間データにおいて圧縮率を向上させるためには非常に有効である。

2.4.2 光線空間符号化に対する要請

光線空間の符号化では、2.4.1節で述べた相関を十分に利用した効率的な圧縮がもちろん必要であるが、以下のような独特な性質も復号化の際に重要となる。

ランダムアクセス性、局所展開性

IBRでは、画像合成時に必要な光線情報が各画像間に分散して配置されている。例えば2.3.1節で述べたLFRでは、ある1本の光線情報を補間するために、隣接の画像から1ピクセル単位で情報を読み出す必要がある。このため、任意の光線情報にすばやくアクセスできることが求められる。

また、ある視点からの画像を合成するには、光線空間の全ての情報を復号する必要はない。ランダムアクセスされたブロックのみを局所的に復号することで、計算コストやメモリなどのリソース消費量を抑えることが望まれる。

プログレッシブ性、スケーラビリティ

画像合成に必要な光線情報の復号には、かなりの時間を要することがある。符号系列にプログレッシブ性を持たせておくことにより、画像合成に必要なデータを全て復号する前に、ある程度品質を落とした画

像を提示することができる。このようなスケーラブルな符号化方式により、ネットワークの帯域やクライアントの計算能力に合った品質の画像の提示を行うなどの応用が考えられる。

2.5 光線空間符号化の具体的手法

2.4節では、光線空間に含まれる相関と、符号化手法に求められる性質について説明した。本節では、光線空間符号化の具体的手法を、以下の3つに分類して説明する。

- 画像群を多次元光線空間データとして均等に扱う手法 (2.5.1節)
- 画像間の予測をブロック探索ベースで行う手法 (2.5.2節)
- 画像間の予測を幾何モデルベースで行う手法 (2.5.3節)

また、それぞれの手法が2.4.2節で述べた性質を備えているかについても議論する。

2.5.1 画像群を多次元光線空間データとして均等に扱う手法

これに分類される手法は、画像内の相関と画像間の相関を同等に扱う手法であり、従来の静止画像圧縮技術を多次元に拡張したものと考えることができる。具体的な手法としては、ベクトル量子化 [7,28] や多次元 DCT [29,30]、多次元ウェーブレット変換 [31,32]、フラクタル符号化 [33] などが利用されている。

ベクトル量子化は、ランダムアクセスを考慮した符号化方式であり、初期の光線空間符号化においてよく利用された。これについては、後に詳細を述べる。

多次元 DCT と多次元ウェーブレット変換は、静止画圧縮技術の標準である JPEG および JPEG 2000 においても利用されている変換符号化方式を、3次元もしくは4次元の光線空間に適用したものである。多次元ウェーブレット変換を利用した手法では、この変換が従来備える解像度・SNR スケーラビリティが実現される。

フラクタル符号化は、画像の自己相似性を利用した符号化方式である。この方式が持つ解像度スケーラビリティに注目し、カメラ視点間の解像度を増加させることにより、光線空間データの補間を行う手法が提案されている [33]。

ベクトル量子化

ベクトル量子化は、複数の光線(画素)をまとめたブロックを、類似した量子化代表ベクトルで表現する符号化である。符号化は次のような2段階の手順で行われる。

(1) コードブックの生成

対象となるベクトル群の一部をトレーニングセットとして取り出し、それらを最もよく近似する代表ベクトル(コードワード)のセットを決定しコードブックとする。コードワードの数によって近似の精度および圧縮率が変動する。

(2) ベクトル量子化

光線空間全体をベクトルに分解し、それぞれのベクトルを最もよく近似するコードワードを選んでいく。

ベクトル量子化後の出力は、量子化代表ベクトルと、元画像中の各ブロックがどの代表ベクトルであらわされるかというインデックス情報である。復号化はコードブックを参照し、インデックス情報をベクトルに置き換えていくだけなので、高速に行うことができる。また、インデックス情報は等長符号化されるので、ブロック単位で光線情報にアクセスすることができ、局所展開性も満たす。

ベクトル量子化では、ランダムアクセス性を重視するために等長符号しか用いていないため、圧縮効率 はあまり高くない。ベクトル量子化後のデータにはまだ冗長性が残っているため、これに対してエントロピー符号化をさらに適用することにより、より高い圧縮率を実現できる。この場合、復号化のスピードとランダムアクセス性は多少犠牲となるが、この後述べる予測を利用した符号化と比較すると、ランダムアクセス性は格段に高い。

2.5.2 画像間の予測をブロック探索ベースで行う手法

画像間の予測をブロック探索ベースで行う手法は、一般に視差補償予測符号化と呼ばれる [34, 35]。従来からの動画圧縮技術である MPEG では、時系列の画像間で動き補償予測が利用されている。一方、多視点画像群においては、時系列画像間の「動き」に似た効果として「視差」が現れるため、視点間の画像に対して MPEG のように予測を利用した符号化を行うことにより、効率的な圧縮が可能となる [36]。

多眼撮像系により取得されるデータにおいては、視点位置の軸 ((2.2) 式における p, q) のデータは画素の軸 ((2.2) 式における θ, ϕ) のデータよりも疎にサンプリングされることが多い。このような場合、視点位置の軸における隣接データ間の相関は少ないため、2.5.1 節で紹介したような、全ての軸に対して変換符号化を行うことは効果的でない。このため、視点位置の軸方向には視差補償予測を行い、残差成分に対して DCT [30] やウェーブレット変換 [37] を行うことが多く検討されている。

視差補償予測を利用した符号化手法では、画像間の相関を効果的に利用することができるため、高い圧縮率を実現できる。一般に、予測の段数を増やすほど圧縮効率は向上するが、それに伴い画像間の依存関係が増えるため、ランダムアクセス性は低下する。以下では、視差補償予測符号化方式の中でも、ランダムアクセス性やスケーラビリティを考慮した手法について説明する。また、動空間への応用についても述べる。

ランダムアクセス性を考慮した視差補償予測符号化

Magnor ら [38] の V-coder では、画像群は MPEG と同じような、I ピクチャと P ピクチャに分けられる。I ピクチャは、DCT を利用してイントラ符号化される。一方 P ピクチャは、近傍の I ピクチャを参照して符号化される。この際、ブロックごとに符号化モードを選択することができ、視差補償を行うかどうか、複数の I ピクチャを参照するかどうか、予測誤差信号を符号化するかどうかを決めることができる。ランダムアクセス性を重視するため、P ピクチャ同士の予測は行われない。

Zhang ら [39-41] も同じように、予測を 1 段階のみに限定することにより視差補償予測による高い圧縮率と、ランダムアクセス性の両立を図っている。この手法においては、2 階層のインデックステーブルを参照することにより復号すべきデータブロックが決定される。

階層的予測符号化

視差補償予測符号化の興味深い応用として、階層的予測符号化を用いた D-coder [38, 42] がある。

D-coder では、図 2.7 に示すように、まず方形の頂点にある 4 つの画像 (A) を符号化する。次に、中心の画像 (B) を A の画像を参照して符号化、さらに A と B を利用して、辺の midpoint の画像 (C) を符号化する。これにより、もとの方形が 4 分割されるので、上記の手順を再帰的に繰り返し符号化を進める。

復号化も符号化と同様の手順で行われるため、光線空間全体がプログレッシブに復号されることになる。つまり、復号化の程度に応じた品質の合成画像を得ることができる。V-coder よりもランダムアクセス性は低下するが、D-coder はスケーラビリティを重視した符号化方式であると言える。

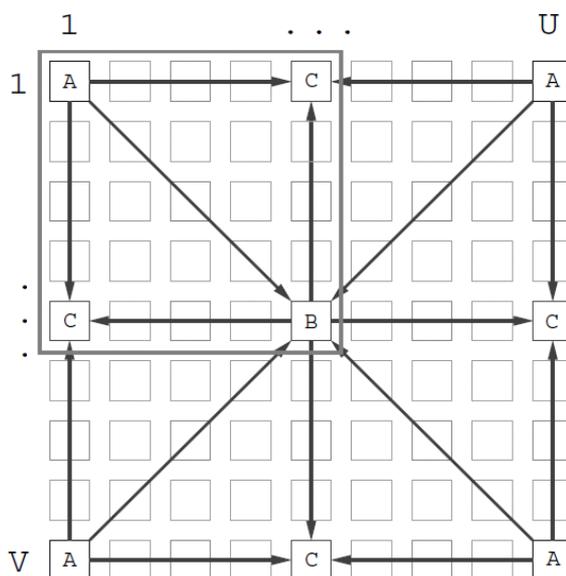


図 2.7: 階層的な多視点画像の符号化順序 (文献 [42] より)

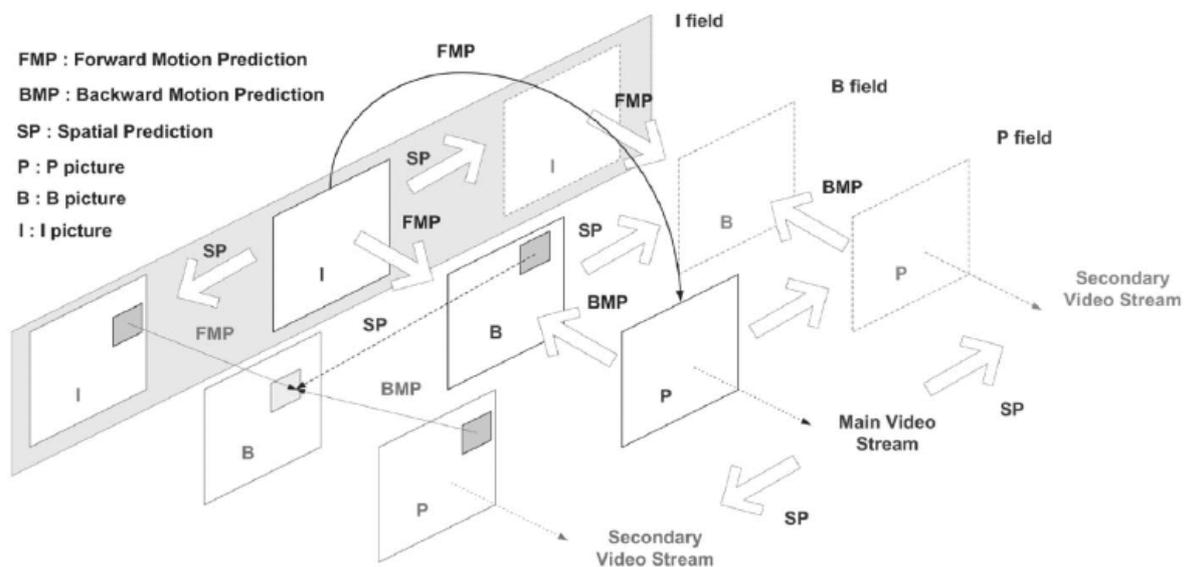


図 2.8: 視差補償予測と動き補償予測を組み合わせた多視点動画の符号化 (文献 [43] より)

視差補償予測符号化の動空間への応用

視差補償と動き補償を組み合わせ、動空間の光線情報を符号化する手法も数多く提案されている [43–46] .

Chan ら [43] は、図 2.8 のように、多眼ビデオシーケンスをメインシーケンスとそれ以外のシーケンスとに分け、符号化を行う手法を提案している。この手法では、メインシーケンスは通常の MPEG-2 を用いて圧縮される。隣接するビデオシーケンスは、メインシーケンスからの視差補償予測および、各々のシーケンス内での動き補償予測を利用して符号化される。Lim ら [44] は、視差補償予測を利用しないメインシーケンスの数を変更することによる、圧縮効率の変化について検討している。これらの手法は、メインストリームを MPEG-2 を利用して符号化しているため、MPEG-2 互換性を持っていると考えることができる。

岡ら [45] は、画像の種類を I ピクチャ、P ピクチャなどと分類せずに、視差、動き補償予測を併用する M ピクチャを全ての画像に対して適用した符号化方式を提案している。

また、木全ら [46] は、GGOP (Group of GOP) 構造を導入することにより、多眼ビデオシーケンスにおけるランダムアクセス性を考慮した符号化手法について検討している。

2.5.3 画像間の予測を幾何モデルベースで行う手法

オブジェクトの幾何モデルを利用することができれば、オブジェクト上の 1 点に対応する光線が各画像においてどこに記録されているかを予測することができる。オブジェクト上の 1 点から発せられた光線間の相関は非常に強く、これを利用することで効率的な情報圧縮を行うことができる。2.5.2 節で説明したブロック探索により画像間の対応点を見つける手法も、幾何モデルを暗に推定していると考えられる。以下では、明示的な幾何モデルを用いたいくつかの符号化手法について紹介する。

仮想オブジェクト面を用いた符号化

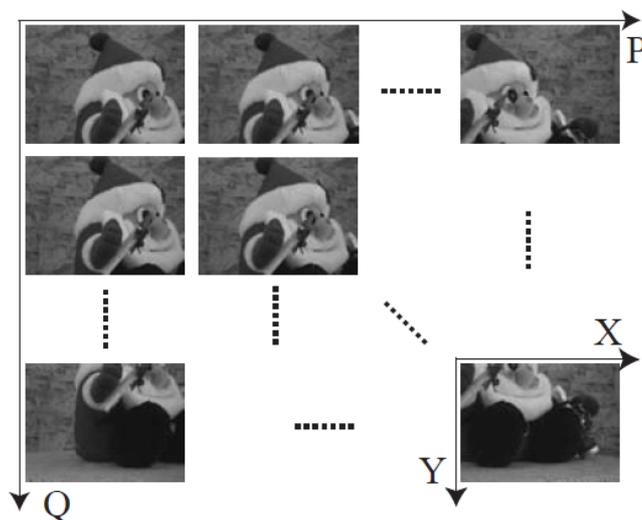
オブジェクトの幾何モデルの最も簡単な近似は多層レイヤモデルである。高野ら [29] は、多層レイヤモデルを仮想オブジェクト面として利用し、効率的な 4 次元 DCT を行う手法を提案している。

まず、オブジェクトの幾何モデルが単一平面モデルで近似される場合を考える。カメラ平面から仮想オブジェクト面までの距離が視差 d ピクセルに対応しているとすると、仮想オブジェクト面上のある 1 点は、一定間隔で撮影された多視点画像群においては、隣接画像間で常に視差 d ピクセルだけずれた位置に撮影される。つまり、図 2.9 (a) のような座標系を多視点画像群に対して定義すると、 Q および Y を固定した PX 平面、いわゆるエピポーラ平面上では、ある 1 点は直線として表現される (図 2.9 (b))。ここで、それぞれの画像を X 軸方向に Nd ピクセルずつずらした空間を考えると、仮想オブジェクト面上の 1 点は P 軸に平行な直線となる (図 2.9 (c))。 QY 平面に対しても同じ変換を施し、その座標系において 4 次元 DCT を実行すると、その係数は P, Q 軸側に極端に偏って分布する。このため、もとの座標系に対して直接 4 次元 DCT を適用するよりも圧縮効率を高めることができる。

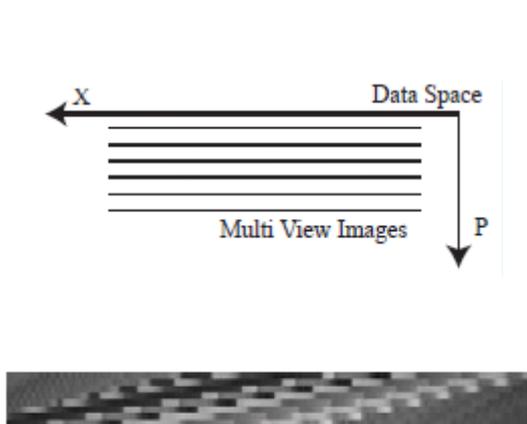
幾何モデルが多層レイヤで近似されている場合には、それぞれのレイヤに属する画像の部分集合に対して別々に上記の変換を行い、4 次元 DCT を用いた符号化を行えばよい。これにより、400:1 程度の圧縮率が実現されている。

Model-Aided / Model-Based Coder

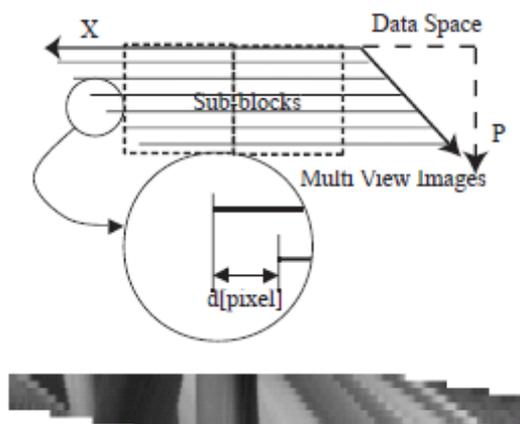
Magnor ら [47–49] は、オブジェクトの幾何モデルを利用した 2 つの符号化方法を提案している。符号化の対象とされるデータは、オブジェクトを取り囲む半球面に沿って撮影された画像群であり、半球面の経度、緯度方向それぞれを均等に分割する位置の 32×8 枚の画像、および半球面天頂からの画像 1 枚からなる。この手法では、符号化の前処理としてオブジェクトの幾何モデルを構築する。多視点画像から 3D メッシュモデルの構築を行うために、Eisert ら [50] のアルゴリズムが利用されている。



(a) 入力多視点画像における座標系



(b) PX平面とエピポーラ平面画像



(c) 仮想オブジェクト面を用いた場合の情報空間

図 2.9: 多視点画像とエピポーラ平面画像 (文献 [29] より)

Model-Aided Coder [47,48] では、複数枚の画像からワーピング予測 [51] を行うことである 1 枚の画像を予測する (図 2.10)．ワーピング予測とは、ある視点から見た画像をモデルの形状に沿って歪ませることで他の視点での画像を予測する手法であり、正方ブロックの平行移動のみで予測を行っている通常の視差補償予測よりも精度の高い予測を行うことができる．また、このような予測を 2.5.2 節で述べた階層的順序で行うことにより、光線空間全体をプログレッシブに符号化する．つまりこの手法は、階層的予測符号化にモデルの概念を導入することで、予測の精度を高めたものと位置づけることができる．

一方 Model-Based Coder [47,49] では、撮影された画像を View-Dependent Texture Map (VDTM) [52] に変換するために幾何モデルが利用される．VDTM とは視点位置に固有のテクスチャマップである．幾何モデル上の 1 点はどの視点の VDTM においても同じ場所にマッピングされるため、VDTM 間の相関はもとの画像よりも高くなる．この VDTM 群に 4 次元ウェーブレット変換を施し、SPIHT (Set Partitioning in Hierarchical Trees) [53] を利用して重要度の順に係数を符号化することにより、高い圧縮率とプログレッシブ

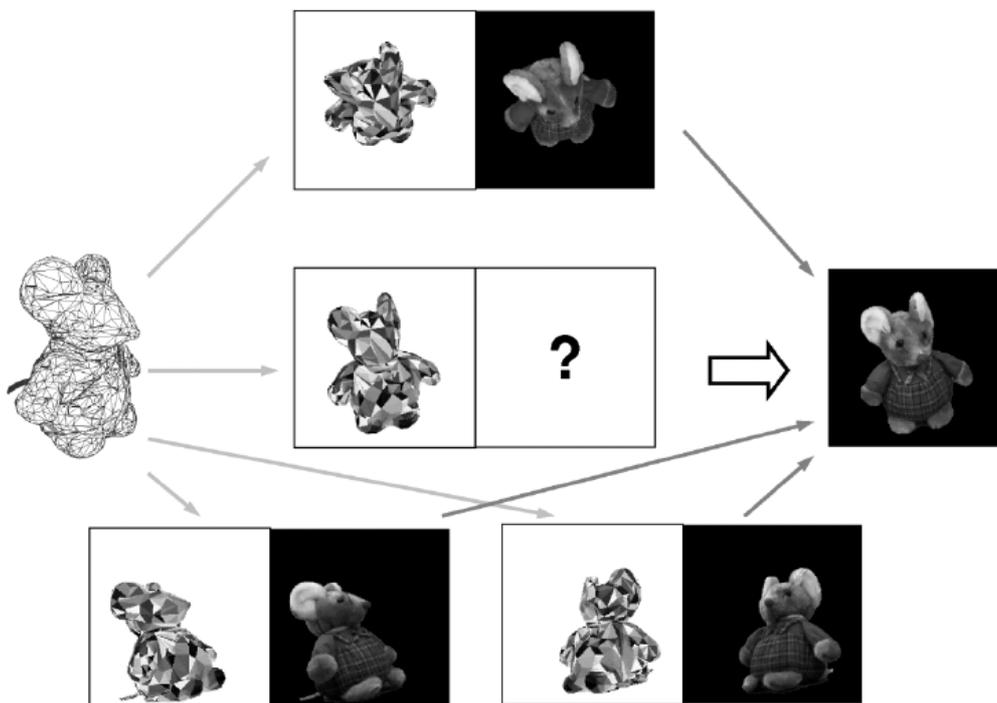


図 2.10: Model-Aided Coder における予測 (文献 [47] より)

ブ性を実現できる .

これら 2 つの手法は , どちらも 2000:1 程度の圧縮率を実現できるが , オブジェクトの幾何モデルが正確な場合は Model-Aided Coder , あまり正確でない場合は Model-Based Coder を利用すると , 同じビットレートで品質の良い合成画像を得られることが示されている . また , Model-Based Coder を動空間に応用する研究も行われている [54] .

Surface Light Field

Surface Light Field (SLF) [55] は , 図 2.11 のように , オブジェクト幾何モデル上の任意の点 u_i における , 任意の方向の光線 ω_j で定義される光線データである . u_i , ω_j に対応する光線 $L_{i,j}$ は Lumisphere と呼ばれ , u_i を中心とした球面上にマッピングできる . SLF の構築は , レンジスキャナなどによりオブジェクトの正確な幾何モデルを取得した後 , 多視点から取得された画像を利用して , オブジェクト上の光線情報を再サンプリングすることにより行われる .

SLF の符号化の際には , まず , 各 u_i におけるメディアン値 \bar{L}_i が求められる . \bar{L}_i は u_i の大まかな色を表しているので diffuse テクスチャと呼ばれ , これを最初に復号することでプログレッシブ性を得ることができる .

$L_{i,j}$ から \bar{L}_i を引いた残差信号の符号化に対しては , ベクトル量子化もしくは主成分分析を応用した 2 つの手法が試みられている . 主成分分析とは , 多次元ベクトル群が持つ情報を少数個の代表ベクトル (主成分) の線形和で表すための統計学的手法である . これは , $\tilde{L}_{i,j} = L_{i,j} - \bar{L}_i$ を ,

$$\tilde{L}_{i,j} \approx \sum_{k=0}^{K-1} c_{i,k} b_{j,k} \tag{2.3}$$

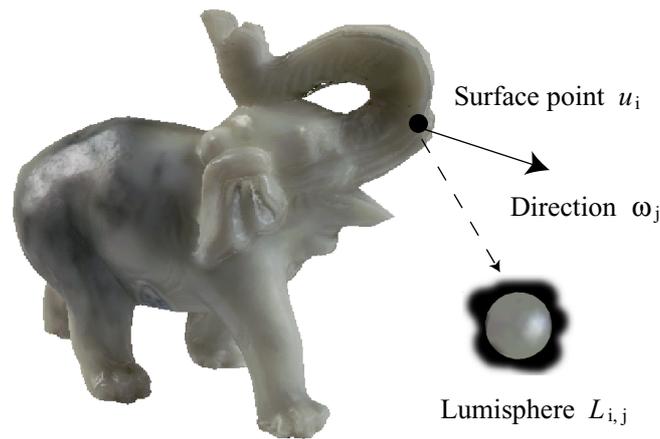


図 2.11: Surface light field (文献 [55] より)

のように分解することに相当し、 $\tilde{L}_{i,j}$ を小さな K で近似することで情報量の圧縮が達成される。

ベクトル量子化と主成分分析とを比較した場合、主成分分析を用いた方がノイズの少ない滑らかな画像を合成できるが、ベクトル量子化を用いた方がオブジェクト表面の反射による視覚的効果が保存されやすいことが示されている [55]。

Chen ら [56] は、(2.3) 式に相当する分解を、Non-negative Matrix Factorization [57] を用いて行うことで、グラフィックスハードウェアの機能を有効に活用する高速なレンダリングを行う手法を提案している。また、分解後の係数に対してさらにベクトル量子化およびグラフィックスハードウェアによるテクスチャ圧縮 (S3TC) を施すことで、5000:1 程度の非常に高い圧縮率を実現している。

Kitahara ら [58] は、分解後の係数に対してウェーブレット変換と SPIHT による符号化を行うことで、プログレッシブ性を保ちながら圧縮率をさらに高める手法を提案している。

2.5.4 符号化手法のまとめ

これまでに述べてきた光線空間の符号化手法をまとめると、表 2.1 のようになる。表 2.1 においては、下に記述されている符号化手法ほどオブジェクトの幾何モデルを積極的に利用している。

画像間の予測を行う手法では、予測の段数を増やすほど高い圧縮率を達成することができるが、それに伴って各画像間の依存性が高まるため、ランダムアクセス性、局所展開性が失われる。高い圧縮効率を実現しつつもインタラクティブな画像合成を行うため、予測の段数を限定するものや、スケーラビリティを持たせることにより、復号の途中段階で品質を落とした画像を提示するといった手法が数多く研究されている。また、高速なレンダリングを行うために、文献 [56] のようにグラフィックスハードウェアの機能を生かせるような符号化を利用する手法も提案されている。

幾何モデルを利用する符号化においては、高い符号化効率を達成できるが、幾何モデル取得の難しさについても考慮しなければならない。例えば SLF では、幾何モデルはレンジスキャナなどを用いて正確に取得する必要があるため、モデルの構築に非常に時間がかかる。Model-Based Coder では、多視点画像から幾何モデルを再構築するため、ある程度速いモデルの取得が可能であり、文献 [54] のように動空間への応用も提案されている。しかし、この手法も完全に自由なシーンに対して適用できるわけではなく、被写体がある 1 つのオブジェクトのみで変化しないという制約が必要である。このように、現段階では幾何モデルを利用した符号化手法は動的なシーンに対してはあまり応用されておらず、従来からの手法であるブロック探索ベースの符号化が多く検討されている。

表 2.1: 光線空間符号化手法のまとめ

References	Geometry	Method	Comp. ratio	Remarks	
Original LFR [7, 28]	no	VQ	Low	Simple random access	
Basic DCP [30, 36, 37]	Implicit	DCP + DCT / wavelet	Middle		
V-coder [38], [39–41]		DCP + DCT	Middle	Considering random access	
D-coder [38, 42]		Hierarchical DCP + DCT	Middle	Progressive decoding	
Multi-view video coding with DCP		[43, 44]	MCP + DCP + DCT	Middle	MPEG-2 compatibility
		[45]			Using M-picture
[46]		Considering random access			
Coding with virtual object surface [29]	Multilayer approx.	4D DCT	High		
Model-aided coder [47, 48]	Reconstructed from multi-views	Warping prediction + Hierarchical DCP	High	Progressive decoding	
Model-based coder		[47, 49]	VDTM + SPIHT 4D wavelet	High	For multi-view videos
	[54]				
SLF coding	Explicit	VQ / PCA	Middle	Progressive decoding	
		NMF + VQ + S3TC	High	Using GPU acceleration	
		PCA + SPIHT 4D wavelet	High	Progressive decoding	

Note: VQ = Vector Quantization, DCP = Disparity Compensation Prediction, MCP = Motion Compensation Prediction, PCA = Principal Component Analysis

また、画像間の予測を利用する符号化方式では、現在のところ、リアルタイムな符号化、復号化は実現されていない。例えば Matusik ら [59] のシステムでは、多眼画像の取得から提示までをリアルタイムで行うため、それぞれのカメラからのビデオシーケンスは独立に MPEG-2 符号化されている。このような場合、カメラの台数に応じて計算機の数を増やすというアプローチを行えるが、画像間の予測を導入した場合、計算機間でデータの受け渡しを行う必要がでてくるため、処理が複雑になり、また、システムの拡張性も乏しいものとなる。効率的な圧縮とリアルタイム性を両立させることも、この分野の今後の課題であると言える。

2.6 まとめ

本章では、光線情報による空間の記述について述べ、実写多視点画像群から自由視点の画像を合成する技術である IBR の基本的な手法を、光線空間の考え方にもとづいて説明した。また、光線空間データの符号化に際して必要とされる要件と、具体的な手法について体系的にまとめた。IBR の実用的な利用の際には、光線空間データの圧縮符号化は必要不可欠なものである。近年では、MPEG の 3DAV アドホックグループにおいても、多眼カメラ動画符号化技術の標準化についての検討が行われている [60–62]。

2.4.2 節で述べたように、光線空間の符号化においては、ランダムアクセス性やプログレッシブ性などの機能が重要視される。2.5 節で述べたような手法では、一般にこれらの機能性は、視点位置に依存せず一様に付加される。次章においては、特にある視点位置における情報に重要度を設定する、階層的な符号化手法を提案する。提案手法においては、重要度を設定された視点の 2 次元合成画像に対してランダムアクセス可能であり、その視点付近の合成画像品質は常に高く保たれる。また、そこから離れた視点での合成画像品質は、ビットレートの向上に伴って改善するという、視点依存スケーラビリティの機能を持つ。

第3章

自由視点画像合成に基づく 光線空間符号化手法の提案

3.1 はじめに

本論文では、自由視点画像合成に基づく光線空間データの視点依存型符号化方式を提案する。これは、ある視点位置およびその周辺の情報を重点的に符号化し、それ以外の情報をビットレートに応じて付加していく、階層的な符号化方式である。本章では、提案符号化手法の目的を述べ、自由視点画像合成に基づいた符号化の枠組みについて提案する。また、提案符号化手法に適した評価方法について検討し、最後に、関連研究に対しての提案手法の位置づけを述べる。

3.2 提案符号化手法の概要と目的

2.5節で述べたような従来の光線空間符号化手法は、一般に、光線空間データを一律に符号化するように設計される。つまり、合成画像品質やランダムアクセスの容易さは、視点位置に関係なくほぼ一定である。また、SNR スケーラビリティの機能性を付加する手法においても、ビットレートに従って全ての視点位置での品質が一律に制御される。

しかし、平面アレイ状に撮影される入力画像群を利用して自由視点画像合成を行うようなシステムを想定した場合、最終的に提示される画像はある視点位置での2次元画像であるため、ある視点位置に対して優先度を設定するような符号化は有用であると言える。低ビットレートにおいては優先度の高い視点位置における2次元画像のみが高品質に合成され、ビットレートの向上とともに、合成画像品質の高い視点数が増大していくような視点依存スケーラビリティが考えられる。

このようなスケーラビリティが活用できるアプリケーション例を図3.1に示す。図3.1(a)のように、ネットワークを介してインタラクティブに画像を合成する場合、伝送すべき最低限の情報は、ユーザの現在の視点位置における合成画像である。しかし、ネットワークは一般に遅延を有するため、現在の視点位置の情報みの送信では、ユーザが視点を変更しようとした場合に往復遅延時間分の遅れが生ずる。このため、現在の視点位置付近の情報も、ビットレートに応じて付加できるような機能性が有用となる。他の例とし

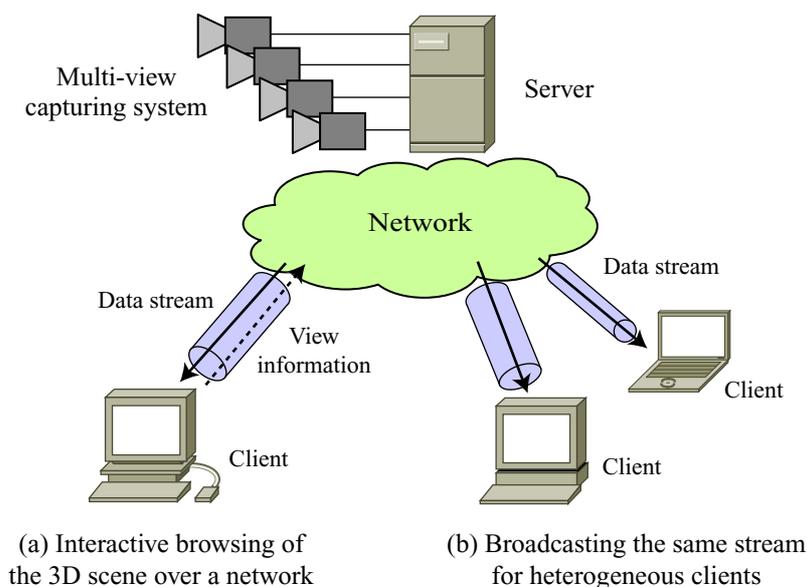


図 3.1: 提案符号化手法が目的とするアプリケーション例

ては、図 3.1 (b) のように、ネットワークの帯域やデータ処理能力の異なる端末へのストリーミングが挙げられる。この場合、帯域の狭い端末には 3 次元空間を代表するような 2 次元画像のみを送信し、帯域が広く処理能力の高い端末に対しては、より多くの視点数を閲覧可能な符号ストリームを送信することが考えられる。

このような目的のもと、本論文では、自由視点画像合成技術を前処理として利用することで、任意の視点位置での 2 次元合成画像を中心とした階層性を実現する、スケーラブルな符号化手法を提案する。提案手法では、まず視域の中から 1 つの視点位置を自由に選び、それを「代表視点位置」と定める。そして、その視点における 2 次元画像をキーとして符号化を行う。これにより、ビットレートに応じた視点依存スケーラビリティを実現する。すなわち、復元可能な視点数と各視点位置における合成画像品質が、ビットレートの向上に伴って改善される機能性を付加する。

3.3 符号化手順

提案手法の符号化手順の概要を図 3.2 に示す。提案手法では、まず入力多視点画像 (a) を利用して、ある与えられた代表視点位置 (b) における合成画像および被写体の視点依存幾何モデル (c) を求める。次に、合成画像と幾何モデルを利用して、各入力画像の予測画像 (d) を合成する。この予測画像と入力画像との差分を取り、予測残差 (e) を求める。以上の手順により求められる、合成画像、視点依存幾何モデル、予測残差を圧縮符号化し、図 3.3 のような階層的符号ストリームを作成する (f)。

(c) における合成画像は、復号時に第一に復元される画像となり、かつ入力画像を予測するための参照画像として利用される。また、視点依存幾何モデルは、鮮明な画像を合成するための副産物として求めるべきものであり、これをリアルタイムに推定して画像合成を行う手法も多く検討されている [12, 13, 26, 63–66]。提案手法では、この幾何モデルを有効利用して符号化効率の向上を図る。

本論文で提案する枠組みは、何らかの方法で与えられた代表視点位置における合成画像を、光線空間符号化の出発点として位置づけるものであり、代表視点位置における合成画像品質は他の視点よりも優先される。代表視点位置としては、被写体全体を俯瞰するような視点位置や、制作者の意図を反映した視点位置を選択することが考えられる。また、図 3.1 (a) のようなインタラクティブな動空間通信を想定した場合には、遠隔地ユーザの現在の視点位置を選択し、その付近の画質を優先的に送信することも考えられる。以下では、代表視点位置に対応する合成画像を「代表視点画像」と呼ぶこととする。

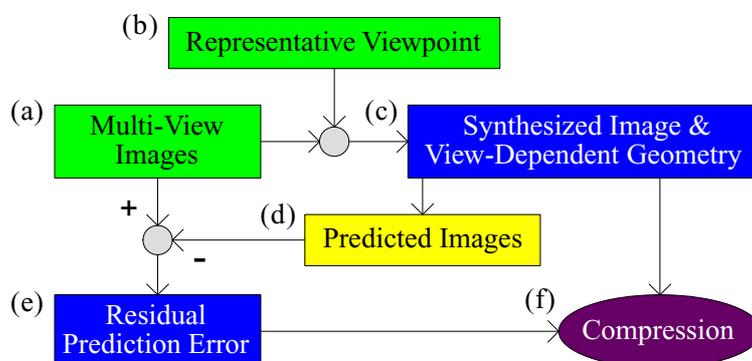


図 3.2: 提案符号化手法のブロック図

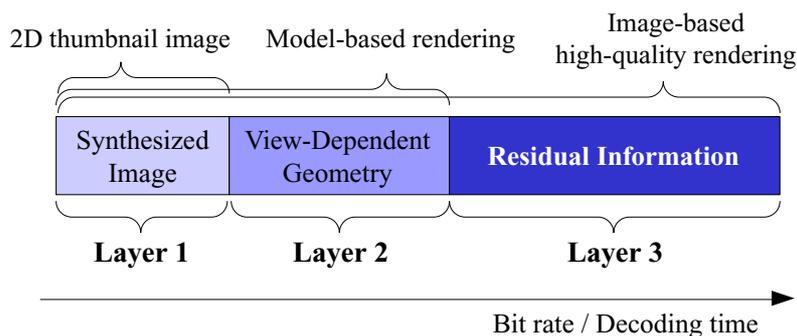


図 3.3: 階層的符号ストリーム構造

3.4 階層的符号ストリーム

提案手法を利用すると、ネットワークの帯域やアプリケーションに応じて、図 3.3 のような階層的な符号ストリームを提供することが可能となる。

レイヤ 1 は 2 次元の代表視点画像である。被写体空間を俯瞰するような代表視点位置を選択した場合、この画像を被写体空間の「サムネイル」として利用できる。レイヤ 2 は視点依存幾何モデルである。レイヤ 1 および 2 のデータを利用することにより、一組のテクスチャとモデルを用いたレンダリング (例えば、文献 [67]) が可能となる。ただし、このようなレンダリングでは、代表視点位置から離れるに従って、オクルージョン領域やモデル精度の影響により、合成画像に劣化が生ずる。レイヤ 3 は残差成分のデータである。これら全てのレイヤを利用すると、全光線空間データによる高品質な画像合成を行うことが可能となる。また、レイヤ 2 までの情報で入力画像群の予測を行うことにより、レイヤ 3 のデータ量を削減し、圧縮効率の向上を図る。

このように、提案手法の符号ストリームでは、各レイヤごとに異なるアプリケーション形態でデータを利用することが可能となる。レイヤ 1 は 2 次元画像であるため、従来の 2 次元画像メディアと互換性を持つフォーマットを作成することも可能である。また、この符号ストリームにおいては、上位のレイヤを利用するほど合成画像品質の高い視点数が増加することに相当する。さらに、レイヤ 3 の残差成分の符号量を制御することにより、代表視点位置付近の合成画像品質は保ちつつ、離れた視点位置での画質をビットレートに従って制御することが可能であると考えられる。次節では、このような機能性を検証するための、合成画像品質の評価手法について述べる。

3.5 提案手法の評価方法

従来の符号化手法は、圧縮後の光線空間から合成される自由視点画像が、視点位置によらず一様な品質を持つように設計されてきた。このため、一般に符号化性能の評価は、復元された入力多視点画像の品質を測定することにより行われる。一方、提案手法は、代表視点位置を中心に据えた手法であり、代表視点位置付近の合成画像品質を優先的に符号化することを目的としている。このような観点から提案手法を評価するには、入力多視点画像よりも自由視点合成画像の品質を評価する必要がある。また、平均的な品質ではなく、視点位置に依存した品質の分布が重要な意味を持つてくる。

そこで、図 3.4 のように、符号化時に利用した代表視点位置 (点 P) を中心として、距離 r を変化させながら合成画像の品質を評価することを考える。距離 r をパラメータとして、圧縮前の光線空間から合成した画像を参照画像とし、圧縮後の光線空間から合成した画像の品質を測定することで、合成画像品質の視

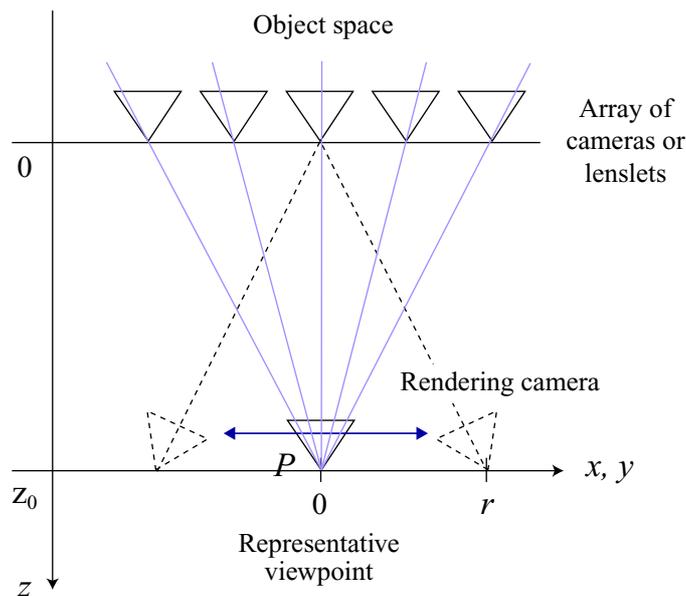


図 3.4: 合成画像品質の視点依存性の評価方法

点位置依存性を検証する。また、残差成分の符号量を変化させた場合、この視点位置依存性がどのように変化するかについても検討することにより、スケーラビリティの評価を行うことができる。

なお、Ramanathan ら [68–70] は、合成画像におけるレート歪特性を最適化するためのデータセグメントの送信方法について検討している。この評価においても、本節で説明した手法と同じように、圧縮前のデータから合成した画像を参照画像として品質の測定が行われている。ただし、Ramanathan らはある視点における合成画像品質を評価しており、視点依存スケーラビリティが評価されているわけではない。

3.6 関連研究との比較

ここでは、本研究の提案手法と、2.5 節で述べたような関連研究との比較を行う。

2.5.4 節でも述べたが、光線空間の符号化においては、一般に画像間予測を多く利用するほど圧縮効率は向上するが、データ間の依存度が高まり、ランダムアクセス性が失われる。ある視点での画像合成の際に必要なデータは、各画像に分散して配置されているため、これを素早く復号するためには、画像間の予測を利用しないベクトル量子化 [7, 28] や、予測の階層を一段階のみに限定する手法 [38–41] などが有利であり、文献 [69] においてもその理論的な枠組みが検討されている。一方、提案手法では、第一に復号する視点位置をあらかじめ決め、その視点での画像を合成するためのデータは画像間予測なしで符号化し、それ以外のデータをモデルを利用した予測により効率的に圧縮する。代表視点位置における 2 次元画像に優先度を与えるアプリケーションを想定し、ランダムアクセスできる視点数を一つに限定した結果、機能性と圧縮効率の両立が達成されることとなる。

また、視点依存スケーラビリティという観点からは、ある一つの入力画像の視点位置を基準視点とし、その視点位置の画像を参照することにより他の視点の画像を符号化する手法が提案されている [42–44, 71]。これらの手法では、基準視点の画像は第一に復号可能であるが、基準視点はあくまで入力多視点画像を撮影した視点位置の中から選択される。一方、本研究の提案手法では、合成画像を求める代表視点位置を視域の中から自由に選択できるため、基準視点の選択の自由度がこれらの手法よりも高い。特に、インテグ

ラルフォトグラフィ入力のように、各視点の入力画像の画角が狭く、1つの視点の画像に被写体のごく一部のみしか写っていないような状況では、被写体全体を俯瞰するような代表視点位置を選択することにより、予測効率が向上すること¹、また、意味のあるサムネイル画像を最初に復元できるという点においても、提案手法は有効であると言える。

3.7 まとめ

本章では、自由視点画像合成に基づく光線空間符号化手法の提案を行った。提案手法では、アプリケーションの利用形態やビットレートに応じて、合成画像品質の高い視点数に対してスケーラビリティを持った符号化を行うことが可能である。また、本章では、提案手法の視点依存スケーラビリティを検証するための評価方法と、関連研究に対する提案手法の位置づけについても述べた。

次章では、提案手法の実装方法の詳細と、階層的符号ストリームを利用した具体的なアプリケーション例について述べる。

¹5.3節, 5.4節でこれを示す。

第4章

カメラアレイ入力および インテグラルフォトグラフィ入力に 対する実装

4.1 はじめに

本章では、第3章で述べた提案符号化手法を、カメラアレイ入力とインテグラルフォトグラフィ入力に対して適用した場合の、具体的な実装例について説明する。まず初めに、符号化器の実装の詳細について述べ、次に、復号器に相当する、提案符号 streams を利用したアプリケーションについて述べる。

4.2 符号化器の実装

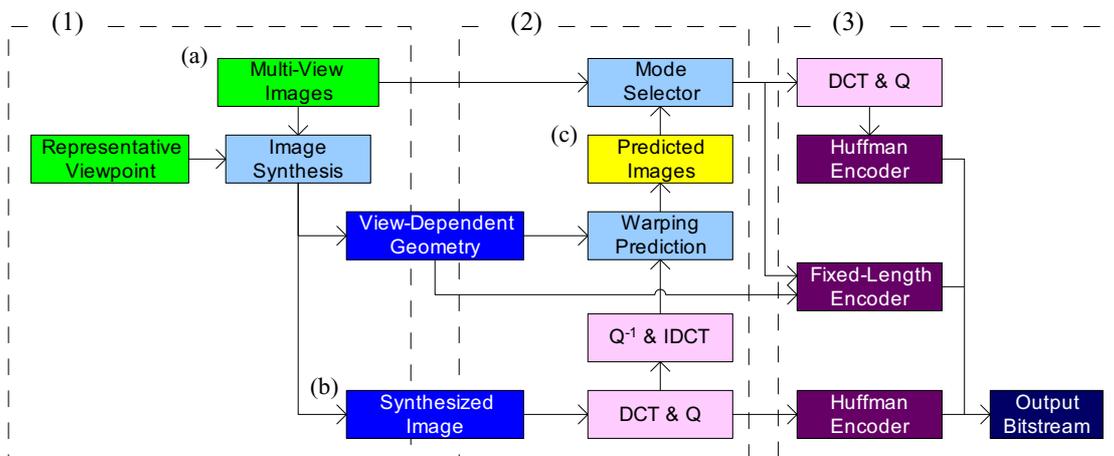
図4.1に符号化器の実装の詳細と、インテグラルフォトグラフィを利用した場合の符号化過程での画像例を示す。以下では、これを3つのステップに分けて説明する。

4.2.1 画像合成と視点依存幾何モデルの構築

提案符号化手法では、まず初めに代表視点位置において、視点依存幾何モデルの構築と代表視点画像の合成を行う。ここでは、カメラアレイ入力、インテグラルフォトグラフィ入力それぞれについて、以下で述べる画像合成法を利用した。

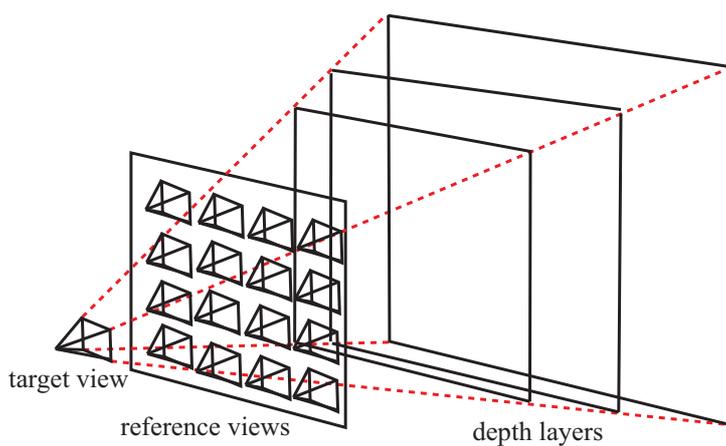
カメラアレイ入力

カメラアレイ入力の画像合成法として、高橋ら [64] の手法を採用した。この手法では、図4.2(a)のように、被写体空間に複数の奥行きレイヤを仮定する。そして、図4.2(b)のように、合成画像における画素(図

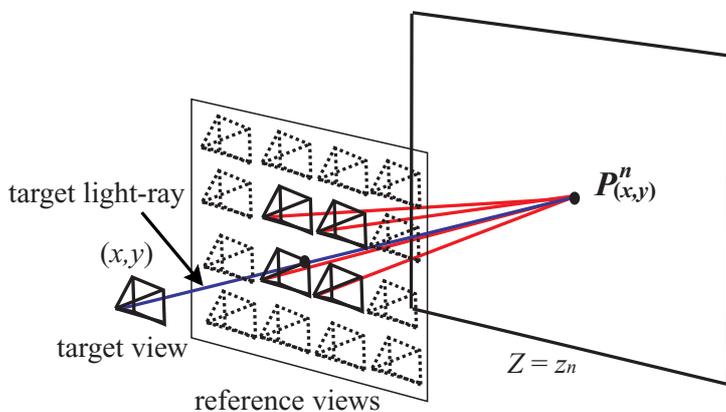


(a) A part of multi-view images (b) A synthesized image (thumbnail) (c) A part of predicted images

図 4.1: 実装の詳細とインテグラルフォトグラフィを利用した画像例



(a) Assumption of depth layers



(b) Evaluation of color-consistency

図 4.2: カメラアレイ入力に対する画像合成法 (文献 [64] より)

4.2 (b) の target light-ray) がそれぞれの奥行きレイヤに属するかを, color-consistency [18, 19] の評価を行うことにより決定する. 評価の際には, ある画素を補間するために利用する近傍カメラの光線のみが利用される. これにより, 合成画像の画素単位での視点依存奥行きマップが求まり, 全ての領域において鮮明な画像が合成される. この手法は, “合焦判定” 法 [26, 27] の一形態であるとみなすこともできる.

インテグラルフォトグラフィ入力

インテグラルフォトグラフィ入力に対しては, Mitsuda ら [63] の手法により画像合成を行った. これは, 図 4.3 のように, 要素画像 1 つに対し 1 つの奥行きを推定し画像合成を行う手法である. 奥行き推定の際にはステレオマッチングを利用し, 要素画像の個数と同じ数の頂点を持つ, 視点依存のポリゴンモデルを構築する. 画像合成は, 推定した奥行き情報を利用して要素画像の一部分を切り出し, ポリゴンモデルにテクスチャマッピングすることにより行われる.

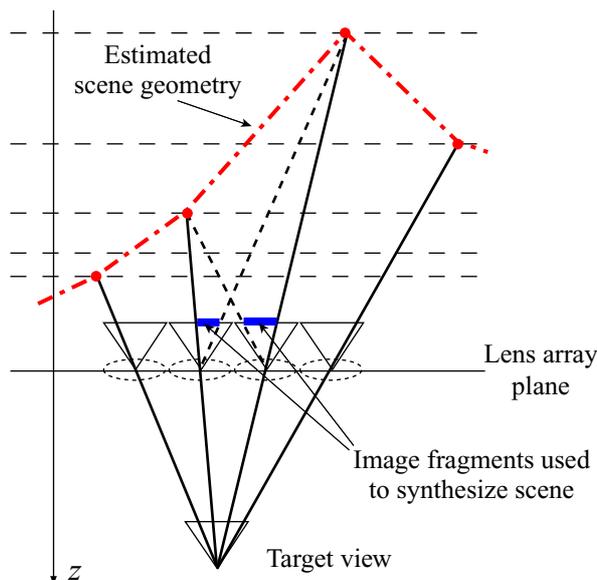


図 4.3: インテグラルフォトグラフィ入力に対する画像合成法

4.2.2 予測画像の合成と符号化モードの選択

以下の処理は、カメラアレイ入力、インテグラルフォトグラフィ入力ともに共通である。

代表視点位置における画像と視点依存幾何モデルを利用して、各入力多視点画像の予測画像を合成する。ここでは、局所復号した代表視点画像を、3Dワーピング [51] することにより予測を行った。予測画像の品質にはばらつきがあるため、 16×16 画素のマクロブロックごとに、符号化モードを以下の3つに分けて符号化した。

- 予測のみ
- 予測および残差符号化
- イントラ符号化

予測画像におけるマクロブロックの画質が、あらかじめ設定した品質 q_{min} より高ければ、そのマクロブロックの付加情報は符号化されない。一方、品質が低い場合には、予測残差ともとの入力画像でそれぞれ分散値を計算し、その値が低い方を符号化する。入力画像が選択された場合には、予測画像は利用されないためイントラ符号化となる。なお、この分散値の比較には、MPEG-2 Test Model 5 [72] で利用されている基準と同等のものを利用した。

4.2.3 階層的ビットストリームの構成

代表視点画像、視点依存幾何モデル、残差情報を圧縮符号化し、階層的符号ストリームに格納する。ここでは、代表視点画像と残差画像に対しては、JPEG を利用した圧縮を行った。幾何モデルとマクロブロックモード情報は、可逆圧縮のために固定長符号化を行った。

なお、4.3 節で述べるアプリケーションでの利用の際には、可逆圧縮する情報を固定長符号としたが、一般的なエントロピー符号化手法を利用することにより、これらの符号量は削減することが可能である。そこで、5 章での符号化性能の評価実験の際には、幾何モデルに対しては、DPCM を行った後のエントロピーを計算し、マクロブロックモード情報に対してもエントロピーを求め、これらを利用してビットレートを

計算した。幾何モデルを構成する点の座標位置は、近傍の点との相関が高いため、DPCMにより効果的な情報圧縮を行うことができる。

4.3 提案符号ストリームを用いたアプリケーション

3.4節で述べたように、提案手法による符号ストリームは、以下のような3つのレイヤで構成される。

- レイヤ1：代表視点画像
- レイヤ2：視点依存幾何モデル
- レイヤ3：残差情報

この階層的符号ストリームは、復号の段階に応じて3つのアプリケーション形態で利用することが可能である。ここでは、それぞれのレイヤを利用したアプリケーション例を示す。

4.3.1 JPEG 互換性を持つ符号ストリーム

提案符号ストリームのレイヤ1は2次元合成画像であるため、従来の2次元画像メディアとの互換性を保った符号ストリームを構成することが可能である。現在普及している2次元画像・動画フォーマットは、一般に、アプリケーション固有のデータを格納することができる拡張領域を持つ。例えばJPEGにおいては、“APP_n” マーカから始まるセグメントがこれに相当する [73]。

そこで、レイヤ1の代表視点画像で通常のJPEGファイルを作成し、レイヤ2およびレイヤ3のデータを表4.1のようにその拡張領域に格納することにより、JPEG互換性を持つ符号ストリームを作成した。この符号ストリームの利用例を図4.4に示す。通常のJPEGビューアでは、拡張領域のデータは無視されるため、代表視点画像のみが閲覧可能である。これは、3次元的なシーンの2次元サムネイルとして利用可能であると考えられる。一方、拡張領域のデータを利用する専用のビューアでは、視点位置を変更しながら画像を合成することが可能となる。

4.3.2 代表視点画像と視点依存幾何モデルを利用したレンダリング

レイヤ1の代表視点画像を、レイヤ2の視点依存幾何モデルにテクスチャマッピングすると、テクスチャ付ポリゴンモデルが作成されることになる。このようなポリゴンモデルを視点位置を変えながらレンダリングすることは、伝統的なコンピュータグラフィックスにおいても多く利用されている。つまり、レイヤ1の2次元合成画像に加えてレイヤ2のデータを利用することで、視点位置の変更が可能になる。

図4.5(a)に、このようなレンダリングを行った場合の画像例を示す。視点位置を自由に変更することは可能であるが、オクルージョン領域においては、代表視点画像には写っていない画素を補間することにより生じる歪が見られる。歪の領域をわかりやすくするために、補間を利用しないポイントレンダリングを行った結果を図4.5(b)に示す。このように、レイヤ1およびレイヤ2のみを利用したレンダリングでは、代表視点位置付近においてはある程度良好な品質が得られるが、そこから離れた場合には大きな劣化が見られる。

一方、図4.5(c)は、レイヤ1からレイヤ3まで全てのデータを利用したレンダリング結果である。画像合成法は、4.2.1節で述べた、視点位置に応じて幾何モデルを推定する手法を用いた。図中で同じ行番号の画像は、同一視点位置でのレンダリングを意味している。全てのレイヤを利用したレンダリングでは、視点位置の変更に伴い、合成に利用される光線が新たに入力画像中から読み出されるため、図4.5(c)のように、どの視点位置においても高品質な画像が合成される。

表 4.1: JPEG 拡張領域に格納されるデータ

Marker	Contents
APP1	Geometry data
APP2	Macroblock mode information
APP3	Elemental image parameters (e.g. width, height)
APP4	Quantization tables
APP5	Huffman tables
APP6	Encoded elemental image data

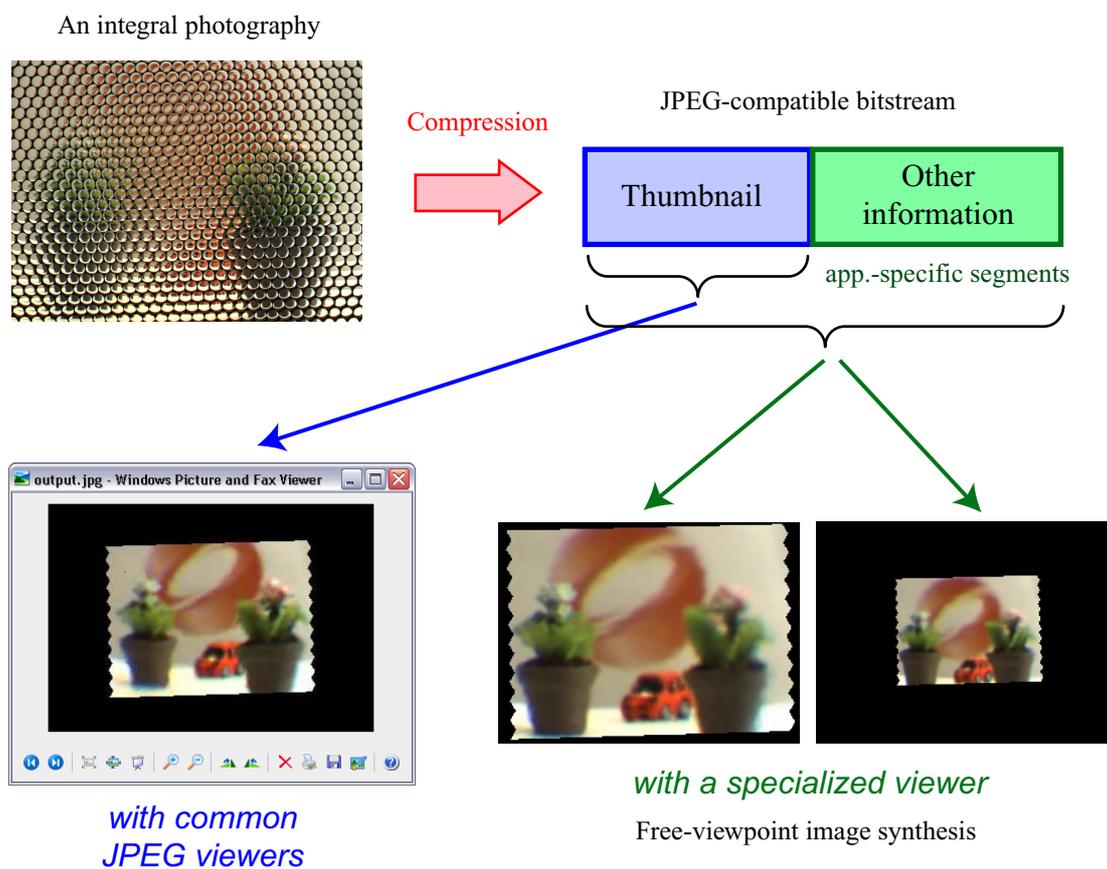


図 4.4: JPEG 互換性を持つ符号ストリームの利用



Thumbnail image (*layer 1*)

Using layers 1 and 2

Using all the layers



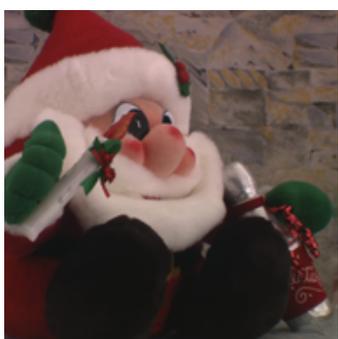
(a) - 1



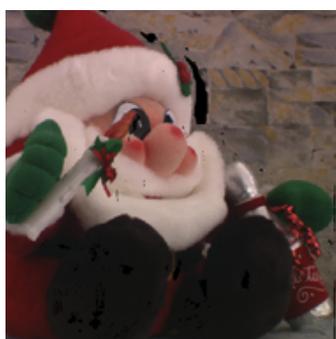
(b) - 1



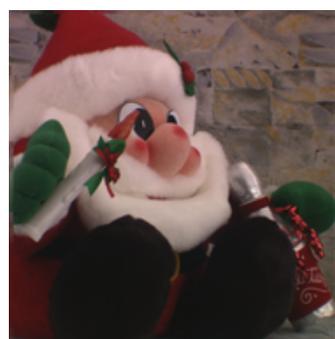
(c) - 1



(a) - 2



(b) - 2



(c) - 2



(a) - 3



(b) - 3



(c) - 3

(a) Texture-mapped polygon

(b) Point rendering

(c) Rendering with depth estimation

図 4.5: カメラアレイ多視点画像を利用したレンダリング例

なお、図 4.5 (a), (b) と図 4.5 (c) とでは、合成画像に写っている範囲が異なっていることがわかる。これは、(a) や (b) のように、ある 1 つのテクスチャ付ポリゴンモデルを利用したレンダリングでは、視点を変更してもレイヤ 1 の合成画像に写っている範囲のみが再現されるのに対し、(c) のような全てのデータを利用したレンダリングでは、入力多視点画像のパラメータにより決定される視域内における画像は全て再現できるためである。ただし、(c) のレンダリングにおいては、カメラ間の光線の内挿のみしか行わないため、外挿となる部分は合成されない。このため、一概にどちらのレンダリング方法が広い視点移動範囲を扱えるかということは難しい。

4.4 まとめ

本章では、提案符号化手法の具体的な実装例について述べた。まず、符号化器の実装の詳細について説明し、次に符号ストリームを復号する段階に応じて、3 つのアプリケーション形態が実現可能であることを示した。代表視点画像と幾何モデルのみを利用したレンダリングでは、代表視点位置から離れた視点位置では、合成画像の品質が劣化することが確認された。

次章では、提案方式の符号化効率および合成画像品質の視点依存性について、より定量的な評価を行う。さらに、残差成分の符号量を変化させた場合の、合成画像品質の視点依存性の変化についての検討も行う。

第5章

符号化効率と 合成画像品質の視点依存性の 評価

5.1 はじめに

本章では、カメラレイ入力およびインテグラルフォトグラフィ入力の多視点画像データを利用して、提案符号化手法の評価を行う。まず、符号化効率を最大化する観点から、代表視点位置を選択する方法について考察する。次に、そのように選択された代表視点位置を利用して、入力多視点画像におけるレート歪特性を評価する。最後に、3.5節で述べたような評価方法を利用して、合成画像品質の視点依存性について検証するとともに、レンダリング方法による合成画像品質の違いについて考察を行う。

5.2 実験条件

本章では、カメラレイ入力とインテグラルフォトグラフィ入力それぞれ2種類のデータを利用して実験を行う。それぞれの入力方式における多視点画像データを、図5.1および図5.2に示す。また、これらのデータの撮影条件を表5.1にまとめる。

カメラレイ画像として、「筑波大学多視点画像データベース」より、81視点で撮影されたサンタクロース人形 (a) Doll)、および街のジオラマ (b) City) の多視点画像群を利用した。各画像は24ビットのカラー画像で、解像度は640×480画素である。

本章の評価実験においては、画像合成に利用する奥行きレイヤの枚数は70とした。代表視点画像の解像度は、512×512画素とした。

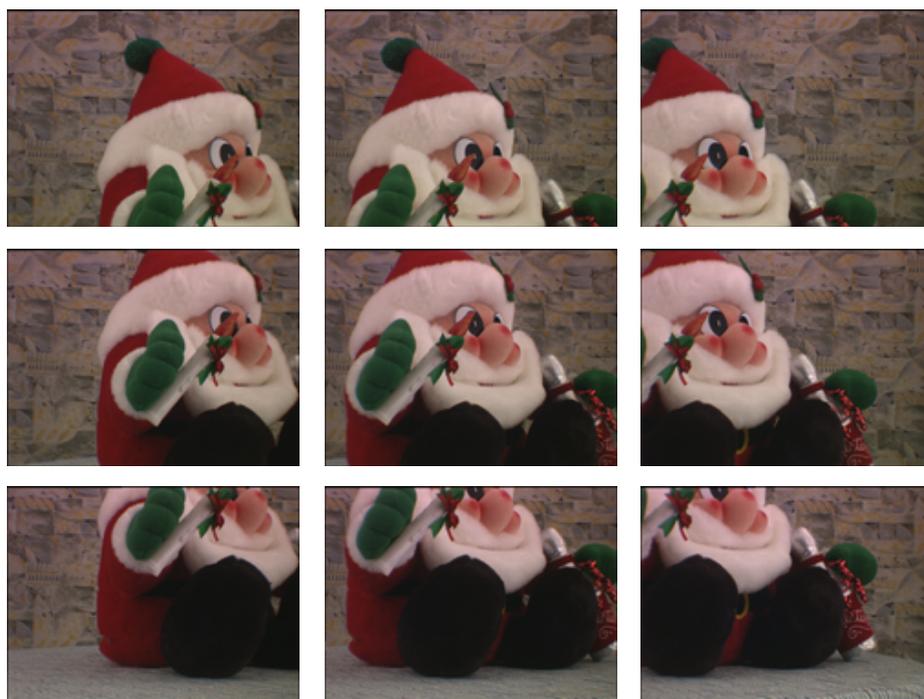
インテグラルフォトグラフィ画像としては、LIFLET [21] により撮影された、玩具の自動車 (c) Toy) と、ミニカーと花 (d) Car and Flowers) のシーンを対象とした。これらは、XGAの解像度を持つ24ビットのカラーカメラ1台でレンズアレイを撮影したものであり、1つの円領域の画像(要素画像)がその視点から見た被写体空間の情報を記録している。ここでは、598個(横26×縦23)の要素画像を31×31画素で切り出し、入力多視点画像群として利用した。なお、円領域の外部は無効領域であり、高周波成分を多く含むため、図5.3のように、円領域内部の画素を利用して水平パディング処理を行うことにより矩形の画像を作成した。

ここでは、画像合成時のステレオマッチングの探索範囲は、1画素精度で10画素とした。つまり、奥行きは10通りに量子化される。代表視点画像の解像度は、ビットレートやPSNRの測定の際には80×56画素とした。ただし、合成画像例の図においては、320×226画素でレンダリングした画像を表示してある¹。

表 5.1: 入力多視点画像群のパラメータ

Name	# of Views	Resolution [pixels]	Arrangement of recording positions	Interval of recording positions [mm]
(a) Doll	81	640 × 480	Regular grid on a plane	20
(b) City	(9 × 9)			8
(c) Toy	598	31 × 31	Honeycomb grid on a plane	1.085
(d) Car and Flowers	(26 × 23)			(pitch of the lens array)

¹これは、前者の画像解像度でも予測に利用するためには十分な情報量を有しており、これ以上解像度を上げて符号化効率は向上しないことが実験的にわかったためである。一方、これを引き伸ばして表示すると解像度の粗さが非常に目立つため、図に表示される画像については、後者の画像サイズで改めてレンダリングを行った。なお、復号後の多視点画像データから自由視点画像を合成する際の解像度は、当然ではあるが代表視点画像の解像度によらず自由に設定できる。

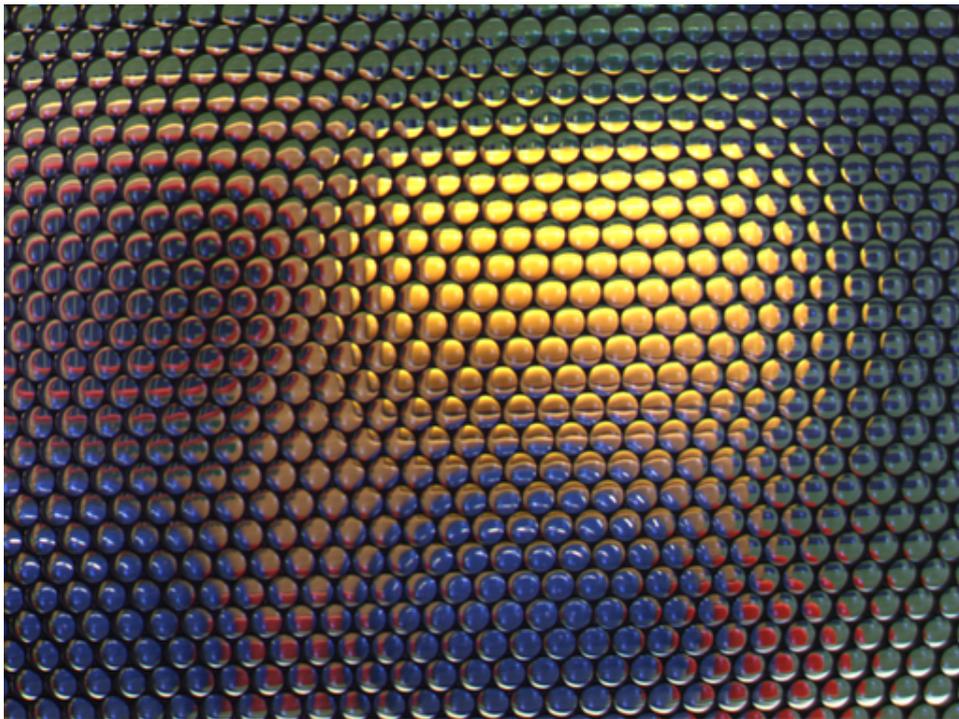


(a) Doll

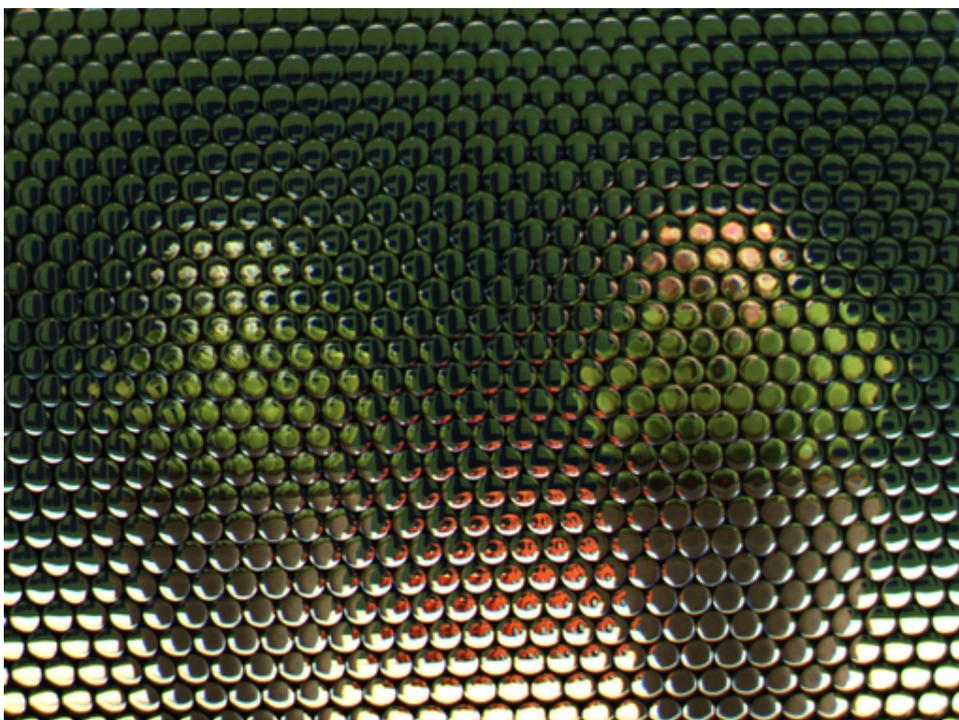


(b) City

図 5.1: カメラアレイ入力多視点画像の一部



(c) Toy



(d) Car and Flowers

図 5.2: インテグラルフォトグラフィ入力画像

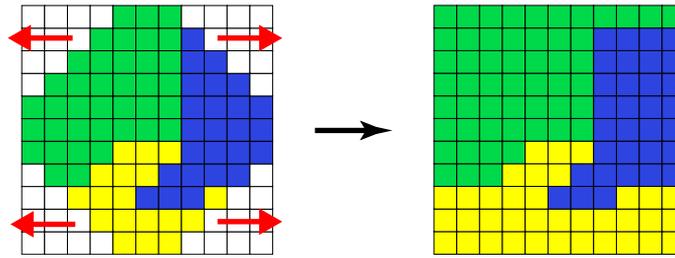


図 5.3: インテグラルフォトグラフィの要素画像に対するパディング処理

5.3 代表視点位置による予測効率の変化

提案符号化手法における代表視点位置の選択基準は、3次元シーンのサムネイルとして適切だと考えられる位置や、遠隔地ユーザの現在の視点位置など、アプリケーションに応じた様々なものが考えられるが、本節においては、符号化効率を最大化する観点から代表視点位置を決定する方法について検討する。

提案手法では、代表視点画像のみを参照し入力画像群の予測を行うため、この予測精度が符号化効率に直接影響を及ぼす。このため、符号化効率を最大化するには、代表視点画像にできるだけ多くのシーンの情報を持たせ、予測効率を向上させることが必要となる。ここで、図 5.4 のように、入力多視点画像の撮影平面の中心を原点とした座標系を考える。入力画像の全てを効率的に符号化するという前提においては、対称性を考慮して、 $(x, y) = (0, 0)$ の z 軸上に代表視点位置を設定した場合に、精度の高い予測のために必要な情報を、代表視点画像に最も多く含ませることができると考えられる。

そこで、視点位置を z 軸上で動かしつつ画像合成と幾何モデルの構築を行い、それらを利用して入力多視点画像を予測した場合の、予測画像の品質を測定した。ある視点位置 z_0 における予測画像の品質は、それぞれの入力画像の視点位置において予測画像を合成し、その輝度値の PSNR 値の平均を求めることにより評価した。なお、レンダリングカメラの解像度は、5.2 節で述べたもので一定とした。 z が大きくなるほど、合成可能な画像の画角が狭くなるため、レンダリングカメラの画角を調節し、できるだけ画面全体に代表視点画像が合成されるようにした。

異なる視点位置において合成される画像例を、(b) City および (c) Toy について図 5.5 に示す。また、このように視点位置を変化させながら、代表視点画像と視点依存幾何モデルから予測画像を合成した場合の、予測画像の品質を図 5.6 に示す。図 5.5 より、視点位置が撮影平面に近すぎる場合 (z が小さい)、シーン全体の一部のみしか合成画像に写らないことがわかる。これは、合成に利用される入力視点数が少なくなるためである。このように z が小さい場合、図 5.6 に示されるとおり、予測画像の品質も悪くなる。一方、視点位置が離れすぎた場合 (z が大きい)、合成画像の画角が狭くなるため、やはり同じようにシーンの一部のみが合成されることとなる。これは特にカメラアレイ入力に対して顕著であり、図 5.6 を見ても、(a) Doll および (b) City の場合は、撮影平面から離れた位置を代表視点位置にすると予測効率が低下することが見てとれる。一方、インテグラルフォトグラフィ入力の場合、もとの入力画像の画角が狭いために、代表視点位置が離れすぎても合成される代表視点画像はあまり変化せず、このため予測効率も極端には減少しない。

以上の考察により、どちらの入力に対しても、シーン全体を俯瞰するような代表視点位置を選択することにより予測効率が最大化できるといえる。つまり、符号化効率の観点からは、代表視点画像を合成する際に全ての入力視点位置における画像が利用でき、かつ、ある程度撮影平面に近い視点位置が、代表視点位置として適していると考えられる。

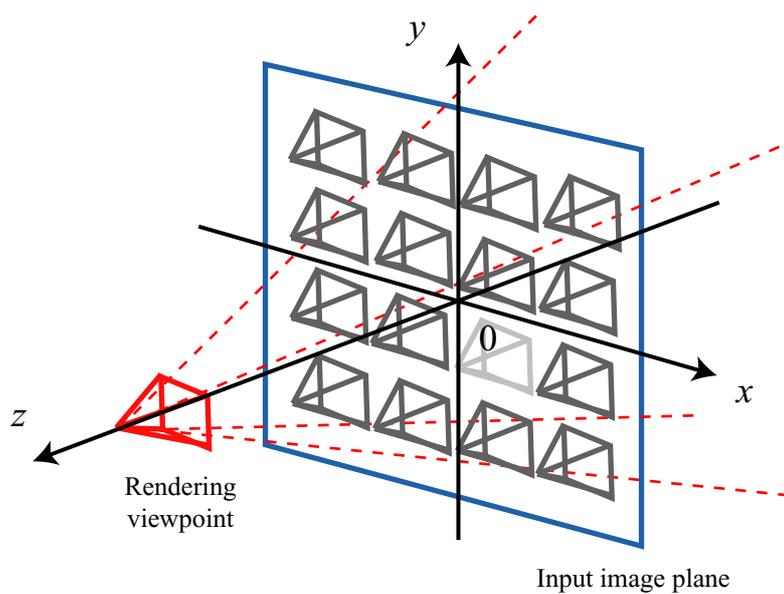
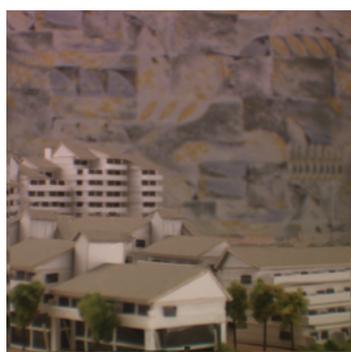


図 5.4: 座標系の設定



$z = 20$ [mm]



$z = 200$ [mm]



$z = 400$ [mm]

(b) City



$z = 20$ [mm]



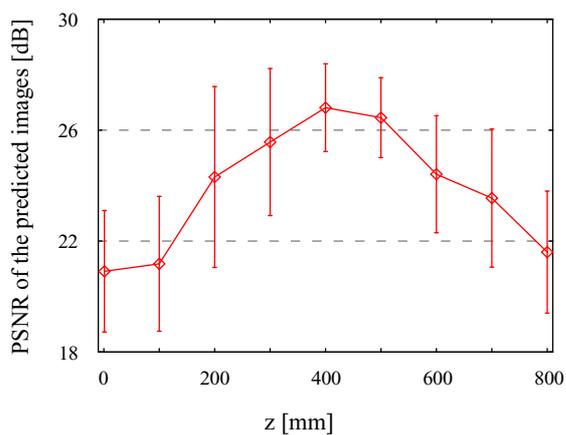
$z = 200$ [mm]



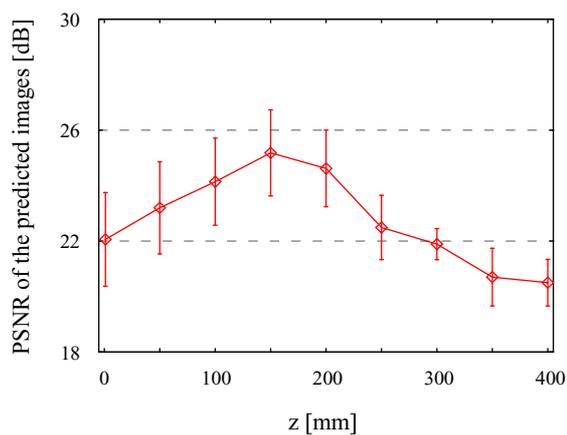
$z = 800$ [mm]

(c) Toy

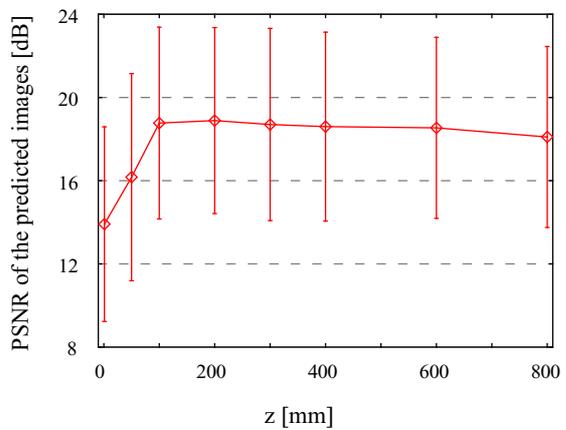
図 5.5: 視点位置による合成画像の違い



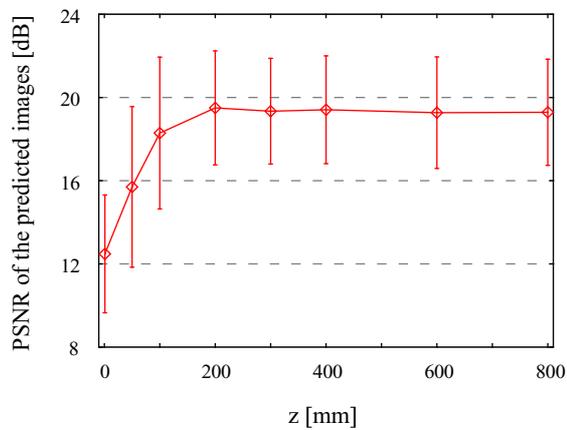
(a) Doll



(b) City



(c) Toy



(d) Car and Flowers

図 5.6: 代表視点位置を前後に動かした場合の予測画像品質の変化

5.4 入力画像群のレート歪特性

5.3 節では、符号化効率を最大化する代表視点位置について検討した。本節では、このような代表視点位置を利用した場合の、入力画像群におけるレート歪特性を評価する。

まず、5.3 節での実験により得られた、予測効率を最大化する代表視点位置を表 5.2 に、その視点位置での合成画像を図 5.7 に示す。提案手法では、この代表視点位置を利用して符号化を行い、その後入力画像群を復号した。このようにして得られた画像それぞれに対して輝度値の PSNR を測定し、その平均値と標準偏差を求めた。また、ここでは代表視点画像と幾何モデルの品質は一定とし、残差成分の品質のみを変化させることによりビットレートを制御した。つまり、残差成分に対するパラメータ q_{min} と、残差画像の JPEG 圧縮の品質を変化させながら測定を行った。

また、比較対象として、全ての入力画像を JPEG 圧縮したものと、画像シーケンスとみなして MPEG-2 圧縮したものの特性を測定した。これらは、予測を利用しないイントラ符号化器、および視差補償予測符号化器の、最も基本的な参照実装であるとみなすことができる。MPEG-2 圧縮の際には、図 5.8 のように、撮影位置に対して横一行の画像をグループとし、各行でスキンの順序を左から右、右から左と切り替えることによって、撮影位置の急激な変化が起こらないようなシーケンスを作成した。また、GOP 構造は $M=3$, $N=15$ 、動きベクトル探索範囲はそのシーケンスにおける最大視差とした。

実験結果を図 5.9 に示す。これより、提案手法は特に低ビットレートにおいて良好な特性を示すことがわかる。予測を行うための付加情報である提案手法での視点依存幾何モデルは、MPEG-2 の動きベクトルよりもオーバーヘッドが少ないため、提案手法では MPEG-2 よりも低ビットレートに到達することが可能である。一方、カメラアレイ入力に対しては、高ビットレートにおいて MPEG-2 の特性が提案手法を上回っている。これは、動きベクトルによる予測の方が、幾何モデルによる予測よりも正確なため、残差成分の情報量が少なくなるためであると推測できる。ただし、MPEG-2 においては視点依存階層性は達成されない。提案手法においても、高ビットレート時には、現在利用している符号化モードに加えて残差同士の予測モードを取り入れることにより、特性が改善されると考えられる。

また、インテグラルフォトグラフィ入力においては、画像間の予測を利用する MPEG-2 よりも、予測を利用しない JPEG の特性が上回っている。このように MPEG-2 の特性があまり良くない原因は、インテグラルフォトグラフィ入力における要素画像は 31×31 画素と非常に小さく、それぞれに被写体のごく一部し

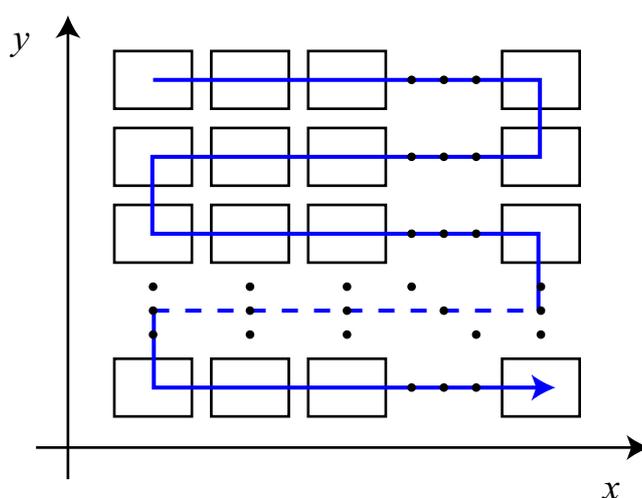


図 5.8: MPEG-2 符号化器におけるスキャン順序

表 5.2: 予測効率を最大化する代表視点位置

Name	(x, y, z) coordinates of the representative viewpoint [mm]
(a) Doll	(0, 0, 400)
(b) City	(0, 0, 150)
(c) Toy	(0, 0, 200)
(d) Car and Flowers	(0, 0, 200)



(a) Doll



(b) City

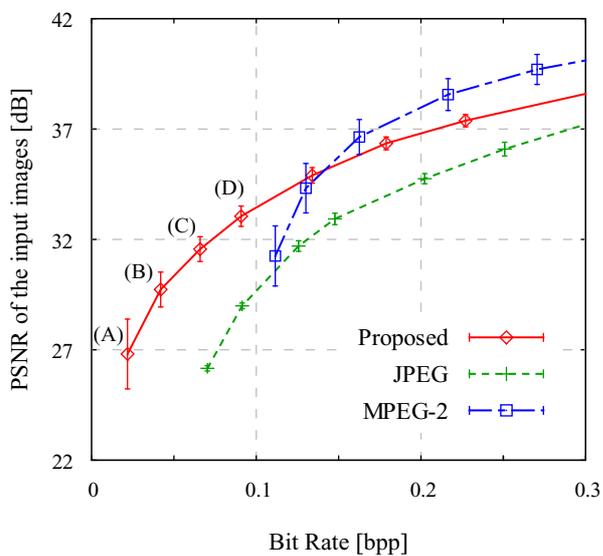


(c) Toy

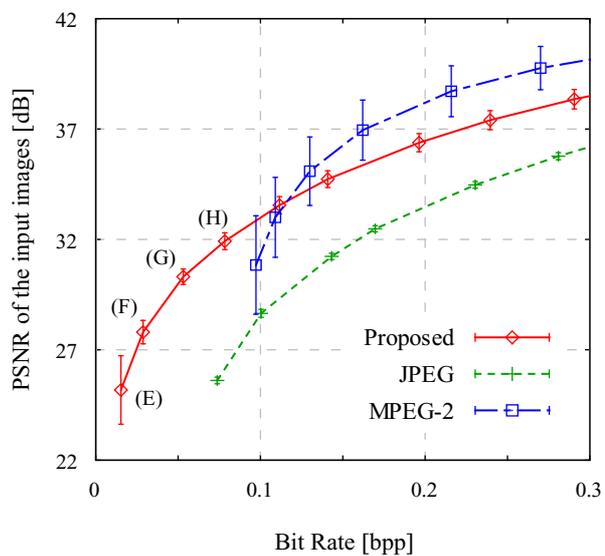


(d) Car and Flowers

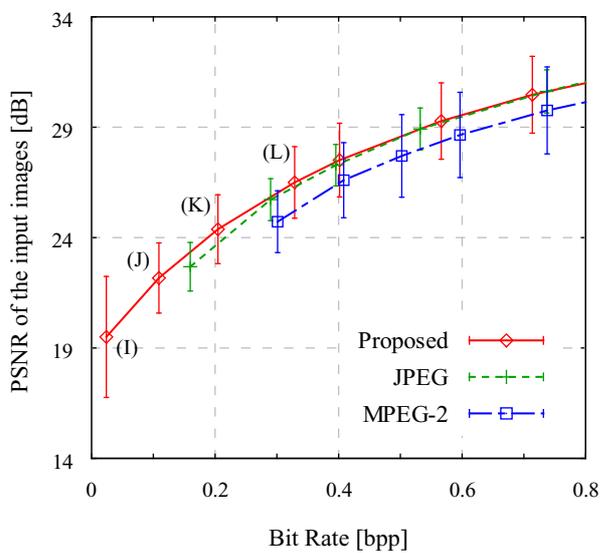
図 5.7: レート歪特性を求める際の代表視点画像



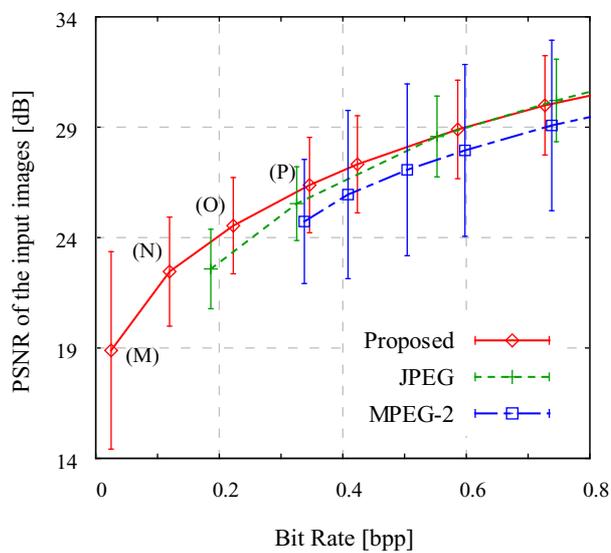
(a) Doll



(b) City



(c) Toy



(d) Car and Flowers

図 5.9: 入力画像群のレート歪特性

か撮影されていないこと、さらに、レンズ歪の影響がそれぞれの要素画像の端に大きく出ているため、要素画像同士の予測は難しいためであると考えられる。一方提案手法では、図 5.7 (c), (d) のように、被写体全体を俯瞰するような合成画像から予測を行ったため、要素画像同士の場合よりも予測を精度良く行えたと考えることができる。

なお、現在の実装では、予測やエントロピー符号化などの基本的なツールとして JPEG や MPEG-2 レベルの技術を利用しているため、これらとの比較実験を行った。H.264/AVC [74] などに用いられている符号化ツールを利用した場合は、提案手法の符号化特性はさらに向上すると考えられるが、その場合比較対象として、入力画像シーケンスを H.264/AVC で符号化したものを選ぶ必要がある。

5.5 合成画像品質の視点依存性

3.2 節で述べたように、提案手法は、2次元画像を中心に据えた視点依存階層性を目的としているため、視点位置に依存した合成画像品質の分布が重要な意味を持つ。そこで本節では、合成画像品質の視点依存性を定量的に評価し、提案手法の有効性を検証する。

5.5.1 評価手法の詳細

ここではまず、評価手法の詳細について述べる。3.5 節で説明したように、合成画像の品質を代表視点位置からの距離に対して測定することにより、合成画像品質の視点依存性を評価する。

図 5.10 に具体的な測定方法を示す。ここでは、視点位置の z 軸方向の位置は代表視点位置と同じ z_0 で

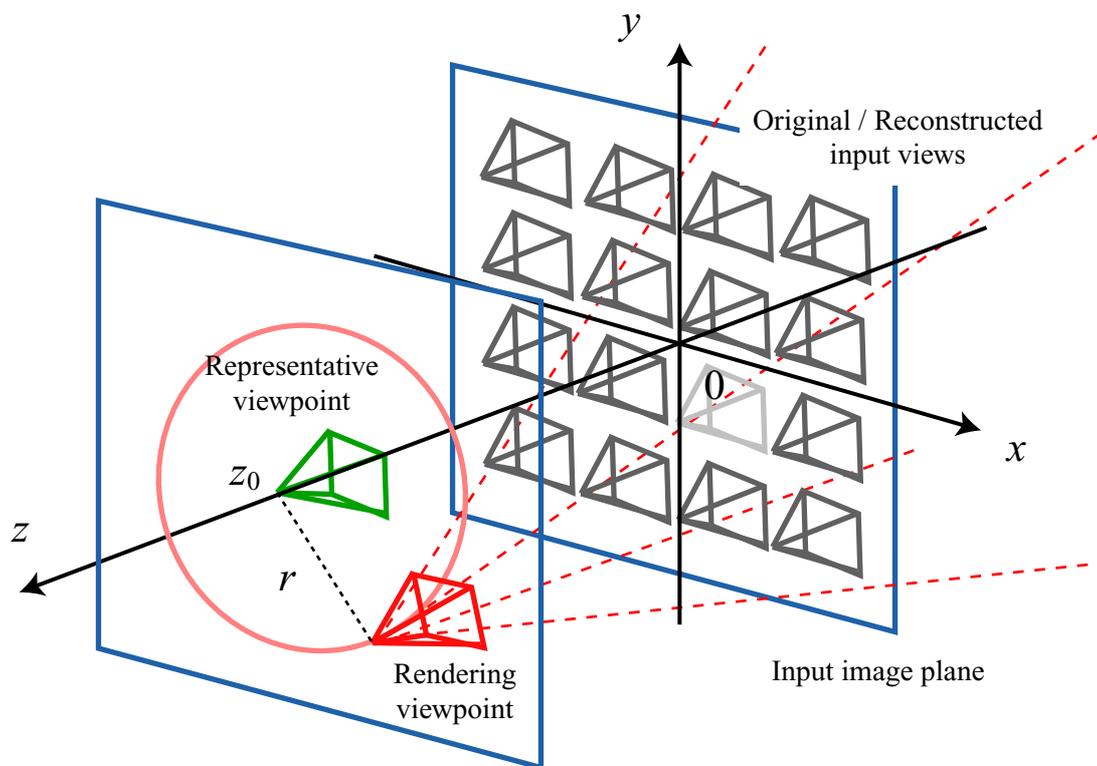


図 5.10: 合成画像品質の視点依存性の評価方法

一定とした上で、半径 r の円周上にて合成画像の品質を測定した。入力画像群をそのまま利用して合成した画像と、圧縮後のデータを復号し入力多視点画像を再構成したものから合成した画像を比較し、輝度値の PSNR 値を求めた。ある半径 r_0 における品質は、円周上に等間隔に配置された 36 箇所の視点位置における、PSNR 値の平均と標準偏差として求めた。また、画像合成法としては、カメラアレイ入力およびインテグラルフォトグラフィ入力それぞれに対して、4.2.1 節で述べたような、符号化の際に用いた手法と同じ手法を利用した。レンダリングカメラの方向は、常に撮影平面の中心、つまり座標軸の原点を向くようにした。

5.4 節での入力画像群におけるレート歪特性の評価と同じように、代表視点画像と幾何モデルの品質は一定とし、残差成分の品質のみを変化させながら上記の測定を繰り返した。これにより、レイヤ 3 のビットレートを増やしていった場合の、視点依存性の変化について評価することができる。また、代表視点位置についても同じように、表 5.2 の通り設定した。

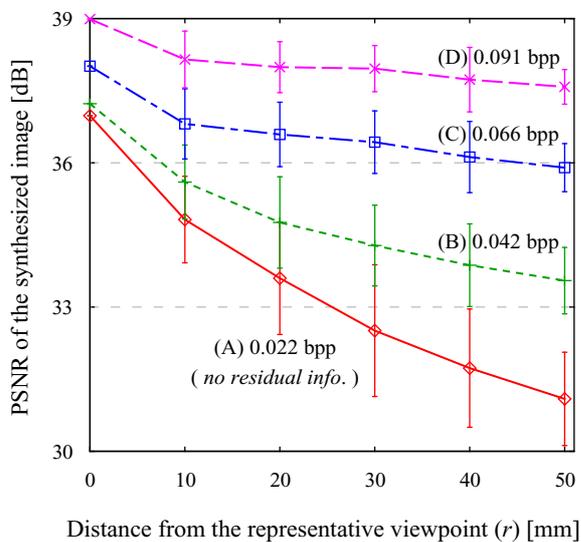
5.5.2 実験結果

図 5.11 に実験結果を示す。図 5.11 の (A) から (P) は、図 5.9 のそれぞれのプロット位置でのビットレートに対応している。(A), (E), (I), および (M) は、残差情報を利用せず、合成画像と幾何モデルのみから入力多視点画像群を復元し、画像を合成した場合の結果である。ビットレートが上がるに従って、利用する残差成分の符号量が増加している。図 5.9 では入力画像全ての品質を評価しているのに対して、図 5.11 では測定の対象となる視点位置における合成画像のみの品質を評価していることに注意されたい。図 5.9 に比べて図 5.11 の PSNR 値が全体的に高くなっているのは、代表視点位置付近ほど画質が良いことを意味している。

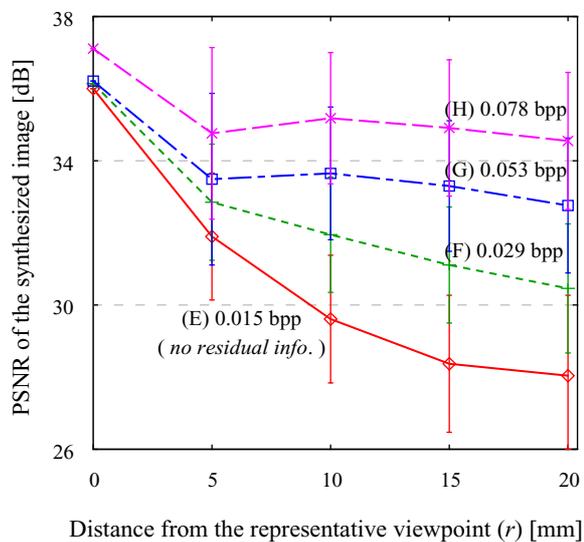
これより、提案手法では合成画像品質に視点依存性が見られることがわかる。残差成分の符号量が少ない低ビットレートにおいては、代表視点位置付近の合成画像品質が高く保たれ、そこから離れるに従って合成画像品質が低下していく。一方、高ビットレートのデータを利用した場合、離れた視点位置での合成画像品質も向上する。提案手法では、残差成分の符号量により、代表視点位置から離れた視点位置での合成画像品質を制御できるといえる。3.4 節で述べたようなレイヤ 1 からレイヤ 3 までの階層性に加えて、この性質を利用することにより、合成画像品質の高い視点数に対してスケーラビリティを持った符号化を実現できるといえる。

なお、(b) City において、 $r = 5$ の視点位置での品質が周辺の視点位置よりも低くなっていることが確認できる。この原因は、図 5.12 のような視点位置において、PSNR 値が平均よりも極端に低くなっているためである。図 5.12 からわかるように、この視点位置においては、入力画像そのものから合成した画像の下部にも劣化が生じている²。この領域は、本来レンダリングの対象としている街のジオラマ以外が撮影されている部分であり、画像合成アルゴリズムにおいて奥行きレイヤを設定する範囲外となっているため大きな劣化が見られる。この領域が PSNR を低下させる大きな要因となっている一方、他の部分においては視覚的な劣化はほとんど見られない。このように、PSNR を計算する際の参照画像として、もとの入力画像から合成した画像を利用しているため、合成アルゴリズムに依存した品質の低下が見られる箇所がある。より正確な測定を行うためには、劣化が起こる部分を検知し（例えば、color-consistency の評価値を尺度とするなど）、その領域を除いて品質の計算をするなどの工夫が必要であると考えられる。

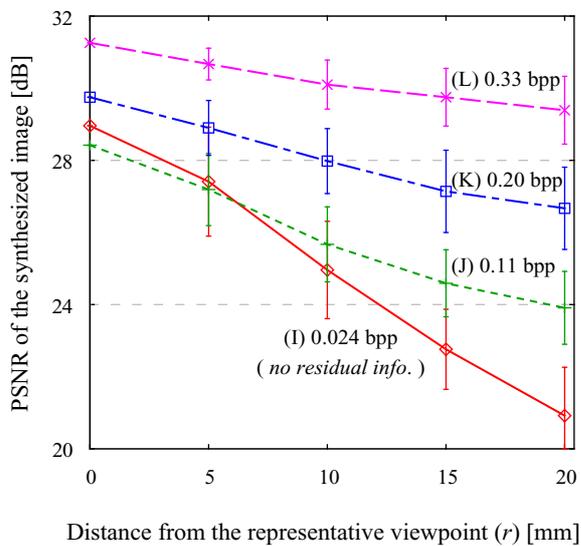
²最下部の黒い帯領域は、入力画像群において光線が取得されていないため、原理的に合成できない領域である。この部分は検出可能であるため、PSNR 計算の際にはもちろん除外されている。ここで言う劣化領域は、最下部より少し上の、光線自体は取得されているが画像合成の際に劣化が生じる領域である。



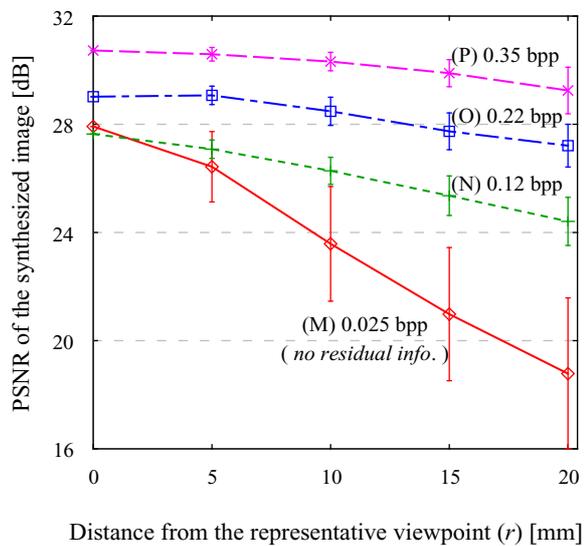
(a) Doll



(b) City



(c) Toy

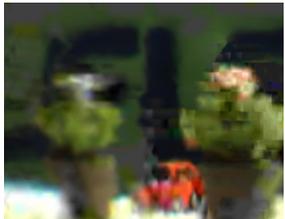


(d) Car and Flowers

図 5.11: 合成画像品質の視点依存性

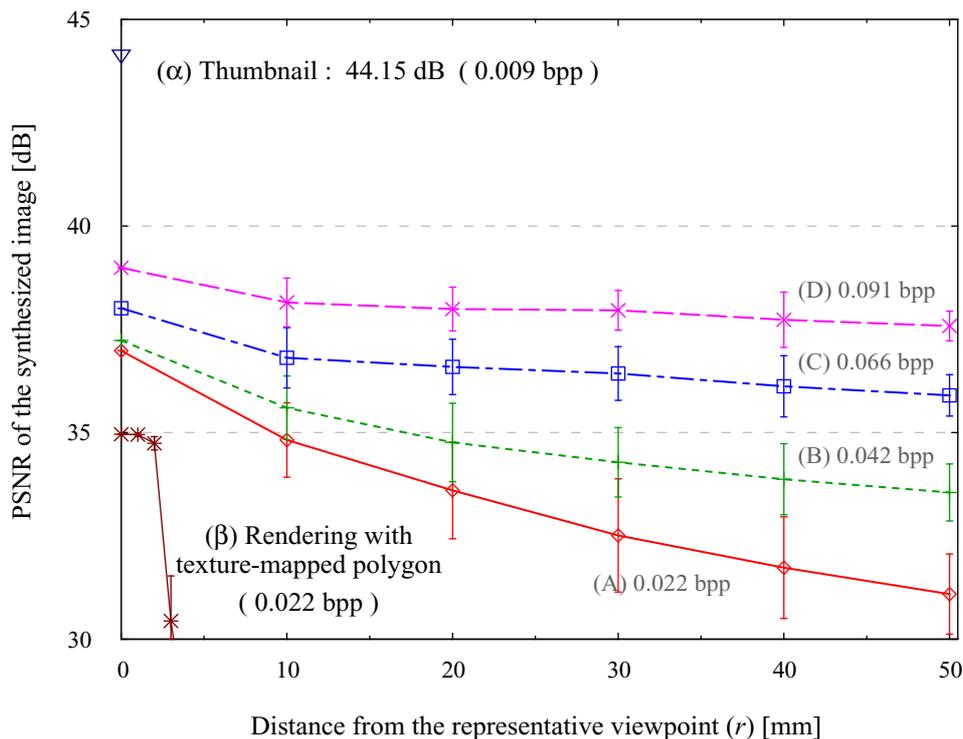
	$(x, y, z) = (0, 0, 400)$	$(x, y, z) = (20, 0, 400)$	$(x, y, z) = (40, 0, 400)$
(B) 0.042 bpp	 PSNR: 37.23 dB	 PSNR: 34.23 dB	 PSNR: 33.11 dB
(D) 0.091 bpp	 PSNR: 38.99 dB	 PSNR: 38.01 dB	 PSNR: 37.46 dB

(a) Doll

	$(x, y, z) = (0, 0, 200)$	$(x, y, z) = (10, 0, 200)$	$(x, y, z) = (20, 0, 200)$
(N) 0.12 bpp	 PSNR: 27.64 dB	 PSNR: 26.92 dB	 PSNR: 24.95 dB
(P) 0.35 bpp	 PSNR: 30.73 dB	 PSNR: 30.38 dB	 PSNR: 29.13 dB

(d) Car and Flowers

図 5.13: 復元される合成画像例



(a) Doll

図 5.14: 各種レンダリング方法を用いた場合の合成画像品質

においても合成画像品質は一定とはなっていない。サムネイルの品質は一定としたため、本来はビットレートによらず品質が一致することが望ましい。

そこで、レンダリング手法を適応的にした場合の合成画像品質について検討する。図 5.14 は、図 5.11 (a) のグラフに、2 つの画像合成手法の特性を追加したものである。(α) はレイヤ 1 のみのデータが利用できる場合を想定したものであり、サムネイル画像の PSNR 値である。レイヤ 1 のみのデータでは視点移動はできないために、代表視点以外では品質を定義することはできない。一方 (β) は、視点依存幾何モデルに代表視点画像をテクスチャマッピングし、これを用いたレンダリングにおける特性である。PSNR を求める際の参照画像は、5.5 節と同じく、符号化の際に用いる画像合成法を入力多視点画像に対して適用した場合の画像である。

これより、代表視点位置のみのサムネイルは、(A) のビットレートの半分以下で復元でき、さらにその品質も高く設定できることがわかる。これからわかるように、実際のアプリケーションでの利用の際には、代表視点位置においてはサムネイル自体を表示すべきである。一方、(β) のレンダリングは代表視点画像と幾何モデルを利用するため、(A) と同じビットレートが必要であり、かつ (A) よりも PSNR 値は劣っている。このように PSNR 値が低い原因は、図 4.5 で示したように、テクスチャマップ付ポリゴンのレンダリングでは、オクルージョン領域や代表視点画像に写っていなかった範囲をうまく再現できないためである。PSNR を求める際には、画素の位置がずれ大きく色が変わることによる影響は極めて大きい。図 4.5 でも示したように、テクスチャマップ付ポリゴンによるレンダリングでも、代表視点位置付近では視覚的には歪がないような画像が復元できるため、この手法は主観的評価よりも低く評価されているといえる。また、(β) のレンダリング手法の方が、(A) における視点移動に応じて幾何モデルの推定を行う手法よりも計算コストは低いため、例えば処理能力の低いクライアントでのデータ利用の際などは、レンダリング手法

としての有用性はあると考えられる。

5.7 まとめ

本章では、提案符号化手法の評価を、カメラアレイ入力およびインテグラルフォトグラフィ入力それぞれ2種類の多視点画像データを利用して行った。まず、符号化効率を最大化する観点においては、シーン全体を俯瞰できるような代表視点位置を選択することが好ましいことを示した。また、このような代表視点位置においてレート歪特性を測定し、一般的な手法よりも符号化効率が向上することを示した。最後に、提案手法では、代表視点位置を中心として合成画像品質が視点依存特性を示すこと、また、残差成分のビットレートにより離れた視点位置での合成画像品質を制御できることを示した。3.4節で述べたようなレイヤ1からレイヤ3の階層性に加え、レイヤ3の残差成分のビットレートを制御することにより、合成画像品質が高い視点数に対してのスケラビリティが実現できると結論付けられる。

第6章

結論

6.1 本研究の成果

本研究の主な成果について述べる。本研究では、自由視点画像合成に基づく3次元空間情報の視点依存型符号化手法を提案した。提案手法では、代表視点位置付近の情報を重点的に符号化し、その視点以外の情報をビットレートに応じて付加するような階層性を実現する。

第3章では、提案手法の目的を説明し、自由視点画像合成に基づく符号化方式の枠組みについて提案した。提案手法の評価方法として、合成画像品質の視点依存性の検証方法について検討した。また、関連研究との比較を行い、その中で提案手法の位置づけについて述べた。

第4章では、カメラレイ入力およびインテグラルフォトグラフィ入力に対する符号化器の実装の詳細について説明した。また、階層的な符号ストリームを利用することにより、復号の段階に応じて3つのアプリケーション形態を実現できることを示した。

第5章では、まず、符号化効率を最大化する観点から、代表視点位置の選択基準について考察し、シーン全体を俯瞰するような視点位置が望ましいことを示した。次に、提案手法の評価を、復元される入力画像、合成画像のそれぞれを利用して行った。復元される入力画像に対しては、レート歪特性を測定し、一般的な符号化手法と比べ、特に低ビットレートにおいて良好な特性が得られることを示した。また、復元される合成画像に対しては、その品質を視点位置に対して測定し、合成画像品質が視点依存性を持つこと、さらに、残差成分を付加していくことにより代表視点位置から離れた視点位置での品質が徐々に向上することを示した。このような評価実験を通じて、提案手法においては、圧縮効率と視点依存階層性という機能を両立できることが示された。

6.2 今後の展望

本研究の今後の展望について述べる。

残差成分における階層的符号化方法の導入

現在の実装では、残差成分の品質を一様に制御しているが、これについても代表視点位置に応じて階層的にすることが考えられる。図6.1に示すように、ある視点位置において画像を合成する際に参照されるのは、入力画像中の一部の領域である。このため、残差成分においても、このような参照領域をROI (Region of Interest) と設定し、優先度を高くして符号化することにより、さらなる視点依存スケーラビリティの向上が期待できる。

代表視点位置の数と選択方法の検討

本論文では、提案手法における代表視点は1つとしたが、これを2つ以上選択するという考えられる。このような場合の符号化効率の変化や、2つ以上のサムネイル画像を利用したアプリケーション（例えば、3次元ディスプレイへの出力に対応する符号ストリーム）なども興味深い。特に、今回の評価実験で利用したような平面上で取得される多視点画像ではなく、被写体を取り囲む半球上で取得される多視点画像群に対しては、1つの代表視点のみでは被写体の裏側などは写らないため、代表視点を増やすことが符号化効率の向上や応用例の拡大につながると考えられる。

また、5.3節においては、符号化効率を最大化する観点から代表視点位置の選択方法を検討したが、実際のアプリケーションにおいては、制作者の意図を反映した視点位置や、遠隔地ユーザの現在の視点位置などを代表視点位置と設定することが考えられる。このような場合、全ての入力画像の圧縮率を向上させるという観点からは、2つ以上の代表視点を選び、そのうちの1つを符号化効率を向上させるために利用す

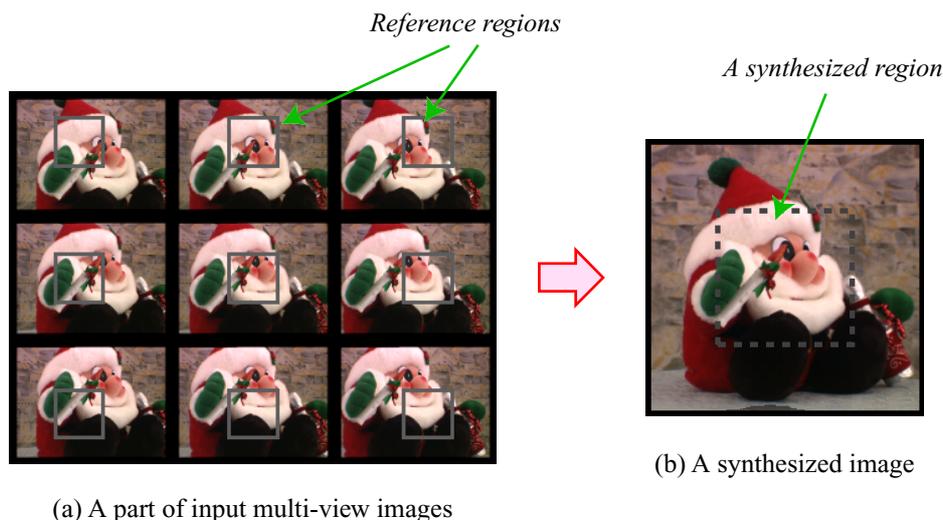


図 6.1: 画像合成に利用される入力画像中の参照領域

るなどといった応用が考えられる．また，前項とも関連するが，代表視点位置とビットレートによっては残差成分の全てを送信する必要のない場合もありえる．

実時間システムへの適用

本論文では静止空間の多視点画像を対象としたが，提案手法のアルゴリズムを動空間に対して拡張し，実時間システムへ適用することも今後の展望としてあげられる．

2.5.4 節でも述べたが，現在のところ，視点間の予測を利用した符号化方式ではリアルタイム性は実現されていない．これは，隣接する視点の画像の参照回数が多く，分散的な処理が難しくなるためである．一方，提案手法においては，代表視点画像を合成する際には集中処理を行う必要があるが，予測画像を求めた後はそれぞれの視点におけるデータの分散処理が可能である．画像合成自体はリアルタイムで行う研究が進められており，また，予測の際には探索を利用しない短時間の処理が可能であるため，全てのプロセスにおいてリアルタイム性を実現できる可能性がある．この場合，残差成分は時系列方向の相関を利用してさらに圧縮することも考えられる．

また，リアルタイムなストリーミングにおいては，先に述べたような ROI 領域の設定が重要となる．ユーザの現在の視点位置での合成画像品質に貢献しない領域の残差情報は送信しないべきであるが，一方，現在の参照領域のみの送信では，視点移動の際に合成画像品質の劣化が見られる．これは，一般にネットワークには遅延があるために生ずる問題である．このような遅延を補償するため，ネットワークの帯域などに応じて ROI 領域を参照領域よりも広く取るといったアプローチを行う必要がある．

現在，筆者の研究室において，図 6.2 のようなカメラアレイ撮像系が構築中である．プロトタイプシステムにおいては，64 台のネットワークカメラ (AXIS 210) がギガビットイーサネットを通じて 1 台の PC に接続されており， 320×240 画素の動画像 (Motion JPEG) が 13 fps 程度で取得・表示されている．このシステムにおいて提案符号化手法を適用，評価していくことが今後の課題といえる．



図 6.2: カメラアレイプロトタイプの外観

謝辞

本研究を進めるにあたり、素晴らしい研究環境と数多くの成果発表の機会を与えて下さり、実に多くの有益なご助言頂きました。原島 博 教授、苗村 健 助教授に心から御礼申し上げます。原島教授には、さまざまな知見や研究への取り組み方についてご教示頂き、研究の方向性について大局的な見地からご指導頂きました。苗村助教授には、研究内容に関する具体的なお助言を頂くとともに、論文投稿や学会発表の際には細かな部分まで親身になって面倒を見て頂きました。本当にありがとうございました。

飯田 誠 助手には、研究内容や発表方法に關して的確なアドバイスを頂くとともに、出張手続きやシステム管理など、研究室での生活における様々な場面で手助けをして頂きました。深く御礼申し上げます。

研究室において、多くの時間を共有させて頂いた SIGLF グループの皆様には特にお世話になりました。博士課程 3 年の高橋 桂太氏には、研究内容や方針に關して実に多くの活発な議論をして頂くとともに、原稿や発表の内容に關して丁寧にご指導頂き、また、研究に対する姿勢に關して多くの見本を見せて頂きました。小島 將氏（現：NHK）には、在学中、レンズアレイ撮像システムに關して多くのことを教えて頂き、研究に關して有益な議論をして頂きました。厚く御礼申し上げます。また、研究はもとより多岐にわたる分野で活発な議論を行い、学会発表の際にも行動をともにした、修士課程 1 年の河 宗玄氏に深く感謝いたします。ともに研究生生活を過ごした、卒論生の金 時煥氏、一松 隆平氏にも感謝いたします。

また、2 年間の研究生生活全般にわたって、原島・苗村研究室の皆様には大変お世話になりました。博士課程 3 年の森村 久美子氏には、研究生生活の中で多くの励ましを頂きました。博士課程 2 年の筧 康明氏、博士課程 1 年の橋田 朋子氏には、発表方法などに關して親切にご指導を頂きました。修士課程 2 年の中洲 俊信氏には、海外発表の際に大変お世話になりました。皆様に心より感謝申し上げます。非常に楽しく充実した学生生活をともに過ごし、お互い切磋琢磨した、修士課程 2 年のチェン・ドミニク氏、西貝 吉晃氏、松岡 薫氏、蓑毛 雄吾氏には、研究のみならずあらゆる場面でお世話になりました。皆様の今後のさらなるご活躍をお祈り申し上げます。また、研究環境の整備に努めてくださった、大谷 智子氏、東山 薫氏、吉田 有里氏に御礼申し上げます。

最後に、研究生生活を通してお世話になりました全ての方々に対して、そしてなにより温かい目で見守って下さいました両親に対して、改めて心から感謝の意を表します。ありがとうございました。

2006 年 2 月 3 日

田口 裕一

参考文献

- [1] C. Zhang and T. Chen: “A survey on image-based rendering — representation, sampling and compression,” *EURASIP Signal Processing: Image Communication*, vol. 19, no. 1, pp. 1–28, Jan. 2004.
- [2] H.-Y. Shum, S. B. Kang, and S.-C. Chan: “Survey of image-based representations and compression techniques,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 11, pp. 1020–1037, Nov. 2003.
- [3] E. H. Adelson and J. R. Bergen: “The plenoptic function and the elements of early vision,” in *Computational Models of Visual Processing* (Eds. by M. Landy and J. A. Movshon), MIT Press, Cambridge, MA, pp. 3–20, 1991.
- [4] 藤井俊彰: “3次元統合画像符号化の基礎検討”, PhD thesis, 東京大学大学院工学系研究科電子工学専攻, Dec. 1994.
- [5] 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”, *テレビ誌*, vol. 50, no. 9, pp. 1312–1318, Sept. 1996.
- [6] 苗村健, 金子正秀, 原島博: “光線情報の正投影表現に基づく3次元空間の記述”, *テレビ誌*, vol. 51, no. 12, pp. 2082–2090, Dec. 1997.
- [7] M. Levoy and P. Hanrahan: “Light field rendering,” in *Proc. ACM SIGGRAPH'96*, pp. 31–42, Aug. 1996.
- [8] A. Isaksen, L. McMillan, and S. J. Gortler: “Dynamically reparameterized light fields,” in *Proc. ACM SIGGRAPH 2000*, pp. 297–306, July 2000.
- [9] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen: “The lumigraph,” in *Proc. ACM SIGGRAPH'96*, pp. 43–54, Aug. 1996.
- [10] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum: “Plenoptic sampling,” in *Proc. ACM SIGGRAPH 2000*, pp. 307–318, July 2000.
- [11] 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, Mar. 2002.
- [12] C. Zhang and T. Chen: “A self-reconfigurable camera array,” in *Proc. 15th Eurographics Symposium on Rendering*, pp. 243–254, June 2004.
- [13] J. C. Yang, M. Everett, C. Buehler, and L. McMillan: “A real-time distributed light field camera,” in *Proc. 13th Eurographics Workshop on Rendering*, pp. 77–85, June 2002.
- [14] B. Wilburn, M. Smulski, K. Lee, and M. A. Horowitz: “The light field video camera,” in *Proc. Media Processors 2002, SPIE Electronic Imaging*, vol. 4674, pp. 29–36, Jan. 2002.

- [15] 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," in *Proc. ACM SIGGRAPH 2005*, pp. 765–776, July 2005.
- [16] M. Tanimoto: "FTV (free viewpoint television) creating ray-based image engineering," in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2005)*, vol. II, pp. 25–28, Oct. 2005.
- [17] 福嶋慶繁, 圓道知博, 藤井俊彰, 谷本正幸: "光線空間法による自由視点画像生成の高速化", 映像メディア処理シンポジウム (IMPS 2005), pp. 57–58, Nov. 2005.
- [18] R. T. Collins: "A space-sweep approach to true multi-image matching," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR'96)*, pp. 358–363, June 1996.
- [19] S. M. Seitz and C. R. Dyer: "Photorealistic scene reconstruction by voxel coloring," in *Proc. IEEE Computer Vision and Pattern Recognition (CVPR'97)*, pp. 1067–1073, June 1997.
- [20] C. Buehler, M. Bosse, L. McMillan, S. J. Gortler, and M. F. Cohen: "Unstructured lumigraph rendering," in *Proc. ACM SIGGRAPH 2001*, pp. 425–432, Aug. 2001.
- [21] T. Yamamoto, M. Kojima, and T. Naemura: "LIFLET: Light field live with thousands of lenslets," *ACM SIGGRAPH 2004 Emerging Technologies*, etech.0130, Aug. 2004.
- [22] M. G. Lippmann: "Epreuves reversibles donnant la sensation du relief," *J. Phys.*, vol. 7, pp. 821–825, 1908.
- [23] 小島將, 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, vol. 59, no. 4, pp. 596–580, Apr. 2005.
- [24] M. Kojima and T. Naemura: "McLiflet: Multiple cameras for light field live with thousands of lenslets," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XII*, vol. 5664, pp. 495–505, Jan. 2005.
- [25] 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, Sept. 2005.
- [26] 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, Mar. 2005.
- [27] K. Takahashi, A. Kubota, and T. Naemura: "A focus measure for light field rendering," in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2004)*, pp. 2475–2478, Oct. 2004.
- [28] H.-Y. Shum and L.-W. He: "Rendering with concentric mosaics," in *Proc. ACM SIGGRAPH'99*, pp. 299–306, Aug. 1999.
- [29] 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, Oct. 1999.
- [30] 山本健詞, 圓道知博, 藤井俊彰, 谷本正幸: "マルチカメラ静止画像の情報圧縮", 画像符号化シンポジウム (PCSJ 2005), pp. 47–48, Nov. 2005.
- [31] M. Magnor, A. Endmann, and B. Girod: "Progressive compression and rendering of light fields," in *Proc. Vision, Modeling, and Visualization (VMV-2000)*, pp. 199–203, Nov. 2000.

- [32] L. Luo, Y. Wu, J. Li, and Y.-Q. Zhang: "Compression of concentric mosaic scenery with alignment and 3D wavelet transform," in *Proc. SPIE Image and Video Communication and Processing*, vol. 3974, pp. 89–100, Jan. 2000.
- [33] 苗村健, 原島博: "自己相似モデリングによる多眼3次元画像の補間と情報圧縮", *テレビ誌*, vol. 48, no. 10, pp. 1215–1221, Oct. 1994.
- [34] M. E. Lukacs: "Predictive coding of multi-viewpoint image sets," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP'86)*, pp. 521–524, 1986.
- [35] T. Naemura, M. Kaneko, and H. Harashima: "Compression and representation of 3-D images," in *Image Processing Technologies: Algorithms, Sensors, and Applications* (Eds. by K. Aizawa, K. Sakaue, and Y. Suenaga), Marcel Dekker, Inc., New York, chapter 4, pp. 90–115, 2004.
- [36] S. Yeom, A. Stern, and B. Javidi: "Compression of 3D color integral images," *Optics Express*, vol. 12, no. 8, pp. 1632–1642, Apr. 2004.
- [37] B. Girod, C.-L. Chang, P. Ramanathan, and X. Zhu: "Light field compression using disparity-compensated lifting," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP 2003)*, vol. IV, pp. 760–763, Apr. 2003.
- [38] M. Magnor and B. Girod: "Data compression for light-field rendering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 338–343, Apr. 2000.
- [39] C. Zhang and J. Li: "Compression of lumigraph with multiple reference frame (MRF) prediction and just-in-time rendering," in *Proc. IEEE Data Compression Conference (DCC'00)*, pp. 253–262, Mar. 2000.
- [40] C. Zhang and J. Li: "Compression and rendering of concentric mosaics with reference block codec (RBC)," in *Proc. SPIE Visual Communication and Image Processing (VCIP 2000)*, vol. 4067, pp. 43–55, June 2000.
- [41] C. Zhang and J. Li: "Interactive browsing of 3D environment over the internet," in *Proc. SPIE Visual Communication and Image Processing (VCIP 2001)*, vol. 4310, pp. 509–520, Jan. 2001.
- [42] M. Magnor and B. Girod: "Hierarchical coding of light fields with disparity maps," in *Proc. IEEE Intl. Conf. Image Processing (ICIP'99)*, pp. 334–338, Oct. 1999.
- [43] S.-C. Chan, K.-T. Ng, Z.-F. Gan, K.-L. Chan, and H.-Y. Shum: "The compression of simplified dynamic light fields," in *Proc. IEEE Intl. Conf. Acoustics, Speech, Signal Processing (ICASSP 2003)*, pp. 653–656, Apr. 2003.
- [44] J. Lim, K. N. Ngan, W. Yang, and K. Sohn: "A multiview sequence CODEC with view scalability," *EURASIP Signal Processing: Image Communication*, vol. 19, no. 3, pp. 239–256, Mar. 2004.
- [45] 岡慎也, 圓道知博, 藤井俊彰, 谷本正幸: "Multi-directional Picture を用いた動的視線空間情報圧縮", *信学技報*, IE2004-94, pp. 15–20, Dec. 2004.
- [46] 木全英明, 北原正樹, 上倉一人, 八島由幸, 藤井俊彰, 谷本正幸: "自由視点映像通信のための低遅延多視点映像符号化", *信学論*, vol. J89-D, no. 1, pp. 40–55, Jan. 2006.
- [47] M. Magnor, P. Ramanathan, and B. Girod: "Multi-view coding for image-based rendering using 3-D scene geometry," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 11, pp. 1092–1106, Nov. 2003.

- [48] M. Magnor, P. Eisert, and B. Girod: "Model-aided coding of multi-viewpoint image data," in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2000)*, pp. 919–922, Sept. 2000.
- [49] M. Magnor and B. Girod: "Model-based coding of multi-viewpoint imagery," in *Proc. SPIE Visual Communication and Image Processing (VCIP 2000)*, vol. 4067, pp. 14–22, June 2000.
- [50] P. Eisert, E. Steinbach, and B. Girod: "Automatic reconstruction of stationary 3-D objects from multiple uncalibrated camera views," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 261–277, Nov. 2000.
- [51] S. E. Chen and L. Williams: "View interpolation for image synthesis," in *Proc. ACM SIGGRAPH'93*, pp. 279–288, Aug. 1993.
- [52] P. E. Debevec, C. J. Taylor, and J. Malik: "Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach," in *Proc. ACM SIGGRAPH'96*, pp. 11–20, Aug. 1996.
- [53] A. Said and W. A. Pearlman: "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, June 1996.
- [54] G. Ziegler, H. P. A. Lensch, N. Ahmed, M. Magnor, and H.-P. Seidel: "Multi-video compression in texture space," in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2004)*, pp. 2467–2470, Oct. 2004.
- [55] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle: "Surface light fields for 3D photography," in *Proc. ACM SIGGRAPH 2000*, pp. 287–296, July 2000.
- [56] W.-C. Chen, J.-Y. Bouguet, M. H. Chu, and R. Grzeszczuk: "Light field mapping: Efficient representation and hardware rendering of surface light fields," in *Proc. ACM SIGGRAPH 2002*, pp. 447–456, July 2002.
- [57] D. D. Lee and H. S. Seung: "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.
- [58] M. Kitahara, H. Kimata, K. Kamikura, and Y. Yashima: "Progressive compression of surface light fields," *NTT Technical Review*, vol. 2, no. 8, pp. 27–34, Aug. 2004.
- [59] W. Matusik and H. Pfister: "3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," in *Proc. ACM SIGGRAPH 2004*, pp. 814–824, Aug. 2004.
- [60] A. Smolic and D. McCutchen: "3DAV exploration of video-based rendering technology in MPEG," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 348–356, Mar. 2004.
- [61] A. Smolic and P. Kauff: "Interactive 3-D video representation and coding technologies," *Proc. IEEE*, vol. 93, no. 1, pp. 98–110, Jan. 2005.
- [62] M. Tanimoto and T. Fujii: "FTV – free viewpoint television," ISO/IEC JTC1/SC29/WG11, M8595, July 2002.
- [63] S. Mitsuda, T. Yamamoto, K. Takahashi, T. Naemura, and H. Harashima: "Interactive view synthesis from integral photography using estimated depth information," in *Proc. SPIE Three-Dimensional TV, Video, and Display II*, vol. 5243, pp. 116–124, Sept. 2003.
- [64] 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 (MIRU 2005), pp. 197–204, July 2005.

- [65] R. Yang, G. Welch, and G. Bishop: “Real-time consensus-based scene reconstruction using commodity graphics hardware,” in *Proc. Pacific Graphics 2002*, pp. 225–235, Oct. 2002.
- [66] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan: “Image-based visual hulls,” in *Proc. ACM SIGGRAPH 2000*, pp. 369–374, July 2000.
- [67] 松村篤志, 内藤整, 川田亮一, 小池淳, 松本修一: “動的に更新される背景バッファを用いた自由視点映像に対する補完方式”, *映情学誌*, vol. 59, no. 6, pp. 901–908, June 2005.
- [68] P. Ramanathan, M. Kalman, and B. Girod: “Rate-distortion optimized streaming of compressed light fields,” in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2003)*, vol. III, pp. 277–280, Sept. 2003.
- [69] P. Ramanathan and B. Girod: “Rate-distortion analysis of random access for compressed light fields,” in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2004)*, pp. 2463–2466, Oct. 2004.
- [70] P. Ramanathan and B. Girod: “Receiver-driven rate-distortion optimized streaming of light fields,” in *Proc. IEEE Intl. Conf. Image Processing (ICIP 2005)*, vol. III, pp. 25–28, Sept. 2005.
- [71] 志水信哉, 北原正樹, 上倉一人, 八島由幸: “奥行き情報を用いた多視点映像符号化に関する一検討”, *画像符号化シンポジウム (PCSJ 2005)*, pp. 45–46, Nov. 2005.
- [72] “MPEG-2 Test Model 5,” <http://www.mpeg.org/MPEG/MSSG/tm5/>.
- [73] ITU-T Rec. T.81: “Information technology — digital compression and coding of continuous-tone still images — requirements and guidelines,” Sept. 1992.
- [74] ITU-T Rec. H.264: “Advanced video coding for generic audiovisual services,” May 2003.

発表文献

- [75] 田口 裕一, 苗村 健: “自由視点画像合成に基づく光線空間符号化”, 映情学誌 (投稿中).
- [76] Yuichi Taguchi and Takeshi Naemura: “View-dependent coding of light fields based on free-viewpoint image synthesis,” submitted to *IEEE Intl. Conf. Image Processing (ICIP 2006)*.
- [77] Yuichi Taguchi and Takeshi Naemura: “Free-viewpoint thumbnail for light field compression,” *ACM SIG-GRAPH 2005 Posters*, #64, July 2005.
- [78] Yuichi Taguchi, Shunsuke Saruwatari, Mikio Hasegawa, Masugi Inoue, Hiroyuki Morikawa, and Tomonori Aoyama: “U¹-Chip: Wireless communication module for fast service discovery,” in *Adjunct Proc. 6th Intl. Conf. Ubiquitous Computing (UbiComp 2004)*, Posters, Sept. 2004.
- [79] 田口 裕一, 苗村 健: “自由視点画像合成に基づく光線空間符号化 — カメラアレイ入力への適用 —”, 画像符号化シンポジウム (PCSJ 2005), P-2.12, pp. 31–32, Nov. 2005.
- [80] 田口 裕一, 苗村 健: “自由視点画像合成に基づく光線空間符号化手法の提案と評価方法の検討”, 第 4 回情報科学技術フォーラム (FIT 2005), vol. 3, J-009, pp. 189–192, Sept. 2005.
- [81] 田口 裕一, 苗村 健: “光線空間符号化のための自由視点画像合成に関する検討”, 3次元画像コンファレンス 2005, 1-3, pp. 9–12, July 2005.
- [82] 田口 裕一, 苗村 健: “自由視点画像合成に基づく光線空間符号化方式の基礎検討”, 信学総大, D-11-144, Mar. 2005.
- [83] 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 回大会, 3A2-2, pp. 475–476, Sept. 2005.
- [84] 金 時煥, 河 宗玄, 田口 裕一, 高橋 桂太, 苗村 健: “自由視点画像合成における見え方の変化を考慮したレンズアレイ撮像系の設計に関する基礎検討”, 信学総大, Mar. 2006 (発表予定).