

NAIST-IS-DD0561004

博士論文

全方位映像のマルチキャストによる  
実時間ネットワークテレプレゼンスに関する研究

石川 智也

2008年3月24日

奈良先端科学技術大学院大学  
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
博士(工学) 授与の要件として提出した博士論文である。

石川 智也

審査委員：

横矢 直和 教授	(主指導教員)
小笠原 司 教授	(副指導教員)
砂原 秀樹 教授	(委員)
山澤 一誠 准教授	(副指導教員)

# 全方位映像のマルチキャストによる 実時間ネットワークテレプレゼンスに関する研究\*

石川 智也

## 内容梗概

遠隔地の情景を高臨場感で提示することでその場に居るかのような感覚を再現する技術はテレプレゼンスと呼ばれる。近年の計算機の高性能化やネットワークの高速化により、環境の撮影から映像提示までを実時間で実行可能な環境が整いつつあり、放送と通信の融合による次世代ネットワークメディアとして実時間ネットワークテレプレゼンスが注目されている。本研究では、遠隔地の情景を自由な視線方向でインタラクティブに観賞可能な全方位映像を用いたテレプレゼンスに焦点を当て、実時間ネットワークテレプレゼンスにおける利用者数の増加に対するスケーラビリティの実現、及び高臨場感なテレプレゼンスのための視線方向のみならず視点位置も自由に変更可能な画像提示技術の実現を目的とする。まず第1章では、テレプレゼンスにおける臨場感実現のための要素技術と、従来研究を概観し、技術的課題を述べる。そして、その課題に対する従来研究での取り組みを述べ、本研究の位置付けと目的を明確にする。第2章では、利用者の増加に対応可能なスケーラビリティを有する実時間ネットワークテレプレゼンスシステムを提案する。提案システムでは、配信する全方位映像は利用者の視線方向に非依存である特徴と、ネットワーク上で同報通信を可能にするマルチキャストプロトコルを利用することでハイスケーラビリティを実現する。第3章では、第2章で提案したシステムを基に自由に視点位置を変更可能な自由視点テレプレゼンスシステムを提案する。多地点に配置した複数の全方位カメラからの画像を用いて実時

---

\* 奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 博士論文, NAIST-IS-DD0561004, 2008年3月24日.

間で自由視点画像を生成する手法を提案し，その手法を用いたプロトタイプシステムを構築する．そして，実験により生成された自由視点画像の定量的画質評価および提案システムの有効性の確認を行う．最後に第4章で，本研究を総括する．

#### キーワード

テレプレゼンス, 全方位映像, ネットワーク, マルチキャスト, 自由視点画像, 実時間処理

# Real-Time Networked Telepresence by Multi-casted Omni-Directional Videos\*

Tomoya Ishikawa

## Abstract

Telepresence is an emerging technology which provides users with a feeling of existing in remote sites by presenting the virtualized real sites. As we can capture images of a scene and display them in real-time, real-time networked telepresence has received much attention as a next generation network media by integrating broadcasting and network communication technologies. This study focuses on the telepresence using omni-directional videos that enables users to see a remote site interactively. The purpose of the study is to realize a high-scalable telepresence system that can support a large number of users and a novel view telepresence system that allows users to change not only the view-direction but also the view-point freely. Firstly, Chapter 1 gives perspectives of the studies for realizing rich telepresence and conventional telepresence systems. Chapter 2 proposes a high-scalable real-time networked telepresence system for supporting a large number of users. To realize the high-scalability, the proposed system takes advantage of the view-independence of omni-directional videos and transfers the omni-directional video using a multi-cast protocol. Chapter 3 then proposes a high-scalable novel view telepresence system based on the system proposed in Chapter 2. Novel view images are generated in real-time from multiple omni-directional images captured at different positions. A prototype of novel view telepresence system is described.

---

\* Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0561004, March 24, 2008.

The quality of generated omni-directional novel view images are evaluated and the validity of the system is successfully demonstrated by the experiments. Finally, Chapter 4 summarizes the present study.

**Keywords:**

telepresence, omni-directional video, network , multi-cast , novel view image, real-time processing

# 目次

1. 序論	1
1.1 実時間ネットワークテレプレゼンスの概観	1
1.2 インタラクティブな映像提示によるテレプレゼンスの従来研究	6
1.2.1 全方位映像によるインタラクティブテレプレゼンス	7
1.2.2 自由視点画像生成によるテレプレゼンス	11
1.3 本研究の位置付けと方針	24
1.4 本論文の構成	28
2. 全方位映像のマルチキャストによる実時間ネットワークテレプレゼンス	29
2.1 序言	29
2.2 全方位画像の平面透視投影変換によるテレプレゼンス方式	30
2.2.1 全方位視覚センサ HyperOmni Vision	30
2.2.2 ハードウェア機能を用いた平面透視投影画像生成	32
2.3 マルチキャストプロトコルによる全方位映像の配信	35
2.4 マルチキャストプロトコルを用いた実時間ネットワークテレプレゼンスシステム	36
2.4.1 システムの設計方針と概要	36
2.4.2 全方位映像コンテンツ	37
2.4.3 Web ブラウザを用いた全方位映像ビューア	41
2.5 ネットワークテレプレゼンス実験	42
2.5.1 高解像度な蓄積型全方位映像を用いたテレプレゼンス	43
2.5.2 車載全方位カメラからの全方位ライブ映像を用いたテレプレゼンス	46
2.6 結言	51
3. 多視点全方位映像を用いたハイスケーラブル自由視点テレプレゼンス	52
3.1 序言	52
3.2 想定するシステムの利用環境	53

3.3	全方位カメラの内部・外部パラメータのキャリブレーション . . . . .	54
3.4	多視点全方位映像からの実時間自由視点画像生成 . . . . .	57
3.4.1	背景差分処理による静的領域と動的領域の分離 . . . . .	59
3.4.2	動的領域の存在領域推定 . . . . .	60
3.4.3	静的領域に対する Morphing による自由視点画像生成 . . . . .	63
3.4.4	動的領域に対する Visual Hull 推定による自由視点画像生成 . . . . .	64
3.4.5	静的領域と動的領域の自由視点画像の統合 . . . . .	67
3.5	シミュレーションによる定量的画質評価に基づくカメラ配置の検討 . . . . .	68
3.5.1	実験目的と条件 . . . . .	68
3.5.2	定量的画質評価結果とカメラ配置に関する考察 . . . . .	73
3.5.3	視差に基づくカメラ配置の定量的評価指標と最適なカメラ配置の推定 . . . . .	86
3.6	ハイスケーラブル自由視点テレプレゼンスシステム . . . . .	90
3.6.1	システムの設計方針と概要 . . . . .	90
3.6.2	サーバ・クライアントの処理内容 . . . . .	91
3.7	複数の利用者による実時間自由視点テレプレゼンス実験 . . . . .	95
3.7.1	システムの構成と実験環境 . . . . .	95
3.7.2	システムの機能実証 . . . . .	96
3.7.3	主観評価による有効性の検証 . . . . .	104
3.8	結言 . . . . .	109
4.	結論 . . . . .	111
4.1	本論文のまとめ . . . . .	111
4.2	今後の検討課題 . . . . .	112
	謝辞 . . . . .	114
	参考文献 . . . . .	115



## 目 次

1	アクティブカメラの例 (左: SONY 製 EVI-D70, 右: Panasonic 製 BB-HCM581)	2
2	アクティブカメラの遠隔操作によるネットワークテレプレゼンスシステム	3
3	全方位映像を用いたネットワークテレプレゼンスシステム (上段: 基本構成, 下段: 複数の利用者への対応)	4
4	姿勢センサ付き HMD による視線追従型テレプレゼンスシステム [OYTY98]	5
5	森田らのシステム構成 [森田 02]	9
6	池田らのテレプレゼンスシステムの外観 [池田 03]	10
7	PanoVi のシステム構成 [YD01]	11
8	視点位置に応じた画像の選択的提示によるシステム (左: Aspen MovieMap [Lip80], 右: Uyttendaele らのシステム [UCK+04])	12
9	レーザーレンジファインダの例 (左: Leica 社製 TCR1105XR, 右: Riegl 社製 LMS-Z360)	14
10	3D Room と 3 次元形状復元した動物体 [SBK03]	15
11	セグメンテーションベースドステレオにより推定された奥行き (左) と自由視点画像例 (右) [ZKU+04]	16
12	View Morphing [SD96]	17
13	Image-Based Visual Hull [MBR+00]	19
14	稲本らの自由視点画像生成手法の概要 [稲本 04]	20
15	稲本らの手法により生成された自由視点画像 [稲本 04]	20
16	ビルボーディング技術を用いた自由視点画像生成 [KKO03]	21
17	平面の集合によって近似的に三次元形状を復元する手法 [KO03]	22
18	環境の 2.5 次元モデルと自由視点画像 [STYY00]	23
19	全方位自由視点画像と平面透視投影画像 [冨手 02]	23
20	HyperOmni Vision の概観 (左) と光学系 (右)	31
21	全方位画像 (右) と平面透視投影画像 (左)	32

22	入力全方位画像（左）と3次元空間中の変換座標（右）	34
23	全方位映像のマルチキャスト配信	36
24	ネットワークテレプレゼンスシステムの概要	38
25	全方位マルチカメラシステム Ladybug	40
26	池田らの手法 [池田 03] により生成された全方位画像	40
27	全方位映像ビューアの構造	42
28	観賞用PCのWebブラウザ画面	45
29	ジャイロセンサを取り付けたHMD	46
30	ライブ型映像のマルチキャスト配信実験のシステム構成	48
31	全方位映像取得車両の外観	49
32	全方位映像取得車両の内部	49
33	利用者それぞれの全方位映像ビューアの画面例	50
34	カメラ配置別の仮想化可能な空間範囲	53
35	システムの利用環境	54
36	世界座標系と推定する全方位カメラの内部・外部パラメータの関係	55
37	環境中のマーカと推定されたカメラの位置・姿勢	57
38	多視点全方位映像から自由視点画像生成処理の流れ	58
39	極座標変換によるヒストグラムの作成	60
40	推定される存在領域	62
41	直線の媒介変数 $M$	62
42	対応点の自由視点画像への投影	64
43	空間直線の Visual Hull の貫通判定	66
44	動物体の存在領域推定による処理の削減	66
45	静的領域と動的領域の自由視点画像の合成	67
46	観光地を想定したシミュレーション環境	71
47	シミュレーション環境中の撮影対象（人物）とカメラ配置（赤丸） 及び自由視点位置（青点）	71
48	基準パターンにおける静的領域の画像とその対応点（上段）およ び動的領域の全方位画像（下段）	72

49	基準パターンにおける PSNR カラーマップ	74
50	生成した自由視点全方位画像（左）と真値の全方位画像（右）	75
51	撮影対象（動物体）とカメラ間の距離を変化させた場合の動物体・カメラの位置関係	76
52	撮影対象（動物体）とカメラ群重心との距離を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：約 0.65m，中段：約 1.3m，下段：約 2.6m）	77
53	カメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合のカメラの位置関係（赤丸：基準パターンと同じカメラ配置，緑丸：カメラ間距離が 1 m のカメラ配置，青丸：カメラ間距離が 4 m のカメラ配置）	78
54	カメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：カメラ間距離 1m，中段：カメラ間距離 2m，下段：カメラ間距離 4m）	79
55	動物体を含まずカメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：カメラ間距離 1m，中段：カメラ間距離 2m，下段：カメラ間距離 4m）	80
56	カメラの解像度を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：512 × 512 画素，中段：768 × 768 画素，下段：1024 × 1024 画素）	82
57	エッジ周辺での画質劣化例（上段：真値の全方位画像，下段：自由視点全方位画像）	83
58	カメラの数を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：3 台，下段：4 台）	84
59	カメラ位置 $P_c$ の拘束（左：カメラを配置不可能な領域への移動，右： $V$ を包含しないカメラ配置への更新）	88
60	室内を模した環境 $G$ ，自由視点位置の移動可能範囲 $V$ とカメラ配置 $P_c$	88

61	非線形最小化処理によるカメラ配置 $P_c$ の変化 (左:初期状態, 中央:繰り返し回数 37 回, 右:繰り返し回数 74 回) . . . . .	89
62	非線形最小化処理による $E_{parallax}$ の変化 . . . . .	89
63	ハイスケーラブル自由視点テレプレゼンスシステムの構成 . . . . .	90
64	プロトタイプシステムの処理の流れ . . . . .	91
65	ライブ映像サーバでのタイムコード付加と自由視点画像生成クライアントでの受信処理 . . . . .	94
66	電磁気センサ (左) と電磁気レーザ付き HMD (右) . . . . .	96
67	実験環境 (左) と推定されたカメラ位置 (右) . . . . .	98
68	自由視点テレプレゼンスにおける遠隔地 (a) と利用者 1 の様子 (b) 及び利用者 1 の位置・姿勢情報を基に生成された自由視点全方位画像 (c) と平面透視投影画像 (d) . . . . .	102
69	自由視点テレプレゼンスにおける遠隔地 (a) と利用者 1 の様子 (b) 及び利用者 1,2 の視界画像 (c, d) . . . . .	103
70	ライブ映像サーバの動作開始後のネットワーク負荷 . . . . .	103
71	複数のライブ映像サーバによる多視点画像配信で発生する時間遅延	104
72	実験時に被験者に提示された映像が取得された環境 (本学ミレニアムホール前) . . . . .	107

## 表 目 次

1	蓄積型映像観賞用 PC の詳細 . . . . .	44
2	画像取得車両のシステム構成 . . . . .	47
3	推定する全方位カメラの内部・外部パラメータ . . . . .	55
4	プロトタイプシステムの機器構成 . . . . .	97
5	自由視点画像生成クライアントでの処理時間の詳細 . . . . .	99
6	12 名の被験者による評価結果 (単位:人) . . . . .	108

# 1. 序論

遠隔地の情景を高臨場感で提示することで、その場に居るかのような感覚を再現する技術はテレプレゼンス (Telepresence) [Moe97] と呼ばれる。テレプレゼンスを最初に提唱したとされる MIT (Massachusetts Institute of Technology) の Marvin Minsky も、ロボットの遠隔操縦において、遠隔地の情報をフィードバックシステムによって利用者に提示することで、そこに居るかのような感覚を与える技術をテレプレゼンスと呼んだ [Min80]。テレプレゼンスとして、システムの利用者に与える感覚は定義されておらず、視覚・聴覚・触覚・味覚・嗅覚のどの情報を用いるシステムにおいても、遠隔にその情報を伝送し再現する技術はテレプレゼンスと呼ばれることが多い。本研究では、それらの感覚の中でも人が外界から得る情報として特に重要とされる視覚情報に焦点を当てる。以下、本論文でのテレプレゼンスとは、視覚情報の提示により時空間的に隔たった場所の感覚を再現する技術を指す。

本章では、実時間ネットワークテレプレゼンスに関する従来研究を概観し、遠隔情景の観賞を目的とするシステムに関しては全方位映像を用いたインタラクティブなテレプレゼンスシステムが適していることを説明する。次に、その全方位映像を用いたインタラクティブなテレプレゼンスシステムの従来研究及び、テレプレゼンスシステムのさらなるインタラクティブ性向上のための自由視点画像生成技術について述べ、自由視点画像生成技術の従来研究を概観する。そして、従来研究の問題点を踏まえ、本研究の位置付けと目的を明確にする。

## 1.1 実時間ネットワークテレプレゼンスの概観

近年の計算機の高性能化やネットワークの高速化に伴い、遠隔地の環境撮影からその映像の提示までを実時間で実行可能な環境が整いつつあり、これまで困難とされていた実時間ネットワークテレプレゼンスの研究が急速に進んでいる [山崎 06]。このような実時間ネットワークテレプレゼンスは、放送と通信の融合による次世代ネットワークメディアとして注目されており、エンターテインメント・遠隔コミュニケーション・医療・教育等様々な分野への応用が期待されてい



図 1 アクティブカメラの例（左：SONY 製 EVI-D70，右：Panasonic 製 BB-HCM581）

る [川西 06, 小野 06, 田中 06] .

それらの応用分野の中でも，観光地や名所の風景をインターネットを介して観賞可能なテレプレゼンスシステム [Par00, Ear96, 北海 99] はとりわけ人気が高く，現在でも広くサービスが行われている．これらのシステムの多くは，遠隔地にパン・チルト・ズーム機構を有するアクティブカメラ（図 1 参照）を設置し，そのカメラの映像をインターネット経由で利用者に送信し提示する．利用者は，単に提示されている映像を観賞するだけでなく，自分が興味のある方向にカメラを向けたり，より詳細に見たい部分をズームアップする等の指示をシステムに与えることができる．従来の TV 放送などでは，放送局から送信されるカメラマンの意図した映像を観賞するのみであったが，ネットワークテレプレゼンスでは，情報の双方向通信が可能となるため，視聴者は自らの希望を反映した映像を観賞することが可能となる．

しかし，これらアクティブカメラの遠隔操作によるテレプレゼンスには，そのシステムの構造上，利用者の要求する映像が利用者に提示されるまでには時間遅延が発生する．その様子を図 2 の模式図に示す．利用者はまず，Web ブラウザ上のコントロールや専用のプログラムにより，見たい方向に対応するカメラの制御信号をシステムに入力する．システムはその制御信号をネットワークを介して，アクティブカメラに伝送する（図中①に対応）．アクティブカメラは伝送された制御信号をもとに，パン・チルト・ズームのパラメータを制御する（図中②に対応）．カメラの制御が完了した後，利用者の要求した映像が映像提示用システム

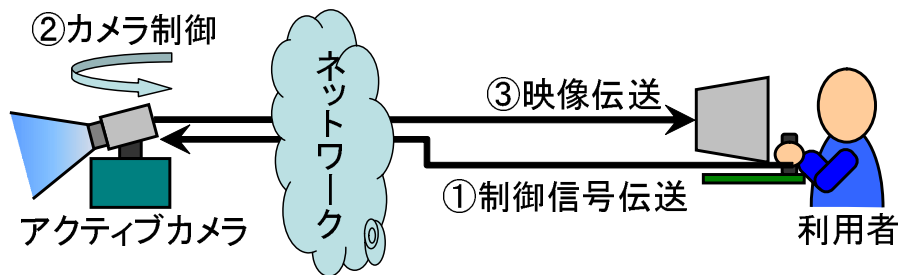


図 2 アクティブカメラの遠隔操作によるネットワークテレプレゼンスシステム

に伝送され、利用者に提示される（図中③に対応）。①～③のそれぞれの処理において時間遅延が発生し、①と③で生じる時間遅延はアクティブカメラと映像提示用システム間の距離やネットワークの状態に依存し、②で生じる時間遅延は、アクティブカメラの性能や目標とするカメラパラメータと現在のカメラパラメータとの差に依存する。極端な例としては、火星探査機にアクティブカメラを取り付け、アクティブカメラを制御しつつ火星の映像を観賞するような場合には、これらの時間遅延は無視できないものとなり、利用者に提示されている映像と制御信号の時間差からシステムの利用は困難となる。

また、このようなテレプレゼンスシステムには操作権の問題がある。アクティブカメラを操作可能な利用者は常に一人であり、複数の利用者がシステムにアクセスした場合には操作権が回ってくるまで待つ必要がある。その間、操作権を取得している利用者以外には各利用者の希望と異なる映像が提示される。

これに対し、信号や映像の時間遅延による影響やシステムを複数の利用者が同時に使用する場合にも操作権の問題が発生しない、全方位映像を用いたテレプレゼンスシステム [エヌ 02, 山澤 02] が提案されている。これらのシステムでは、周囲 360 度の情景を一度に撮影することが可能な全方位カメラ [COH85, YNY91, 山澤 96, KYI+98, 棚橋 02, 池田 03] により全方位映像を取得するため、アクティブカメラのようにパンやチルトの操作を行う必要がなく、全方位映像の一部を切り出し、変換することで利用者の要求する視界画像を提示することが可能となる。

全方位映像を用いたテレプレゼンスシステムの基本的な構成を図 3 上段に示す。遠隔の全方位カメラは撮影した全方位映像を画像変換用 PC に送り続け、画像変

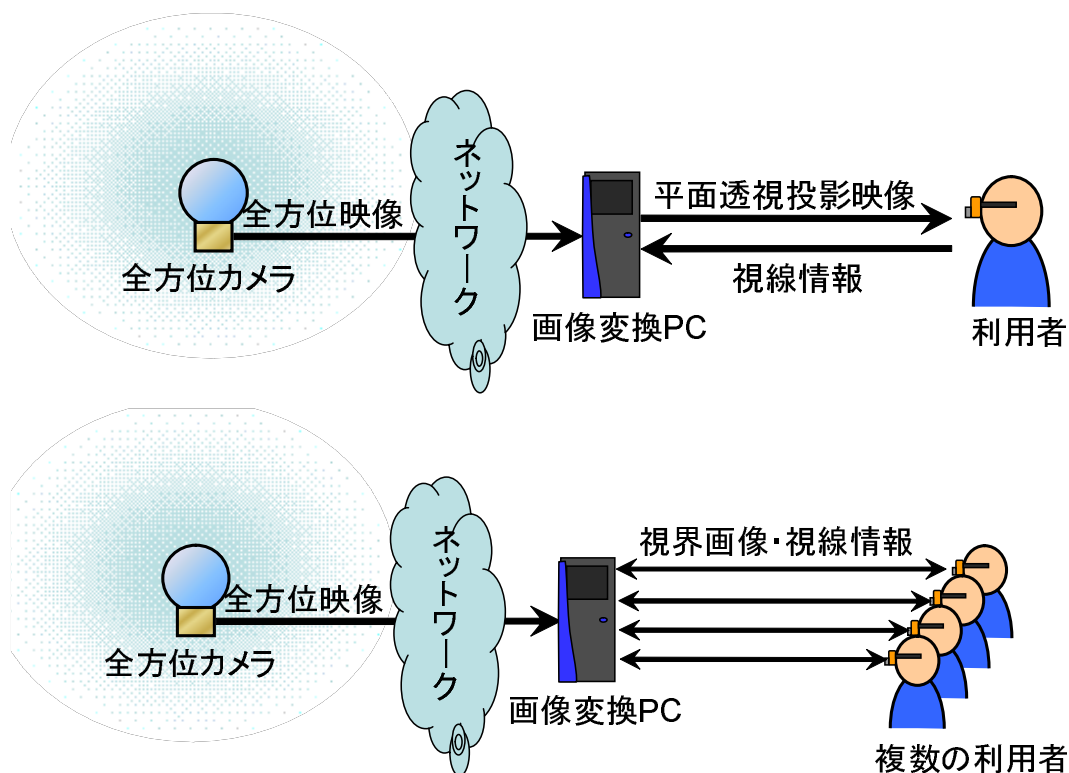


図3 全方位映像を用いたネットワークテレプレゼンスシステム  
 (上段：基本構成，下段：複数の利用者への対応)

換用PCは受信した全方位映像と利用者からの視線情報を基に一般的なカメラで撮影したような平面透視投影画像を生成し利用者に提示する。全方位カメラによる撮影と利用者に対する映像提示の間には距離に依存する時間遅延が発生するが、その遅延は利用者の視線情報に対応した視界画像の生成には影響しないため、見直しに関しては時間遅延を殆ど感じないインタラクティブな観賞を行うことができる。また、全方位画像から平面透視投影画像への変換処理を高速に行うことで、複数の利用者それぞれの要求する視界画像を提示可能である(図3下段参照)。

このような全方位映像を用いて任意の視線方向の画像を瞬時に生成し提示するテレプレゼンスは、利用者の動作に追従した映像提示を行うことで高いインタラクティブ性と臨場感を実現している [OYTY98, 森田02]。これらのシステムでは、利用者はHMD(Head-Mounted Display)と頭部の姿勢を計測するためのセン





図 4 姿勢センサ付き HMD による視線追従型テレプレゼンスシステム [OYTY98]

サを装着し，センサからの姿勢情報を基に生成された視界画像を觀賞する（図 4 参照）．頭部の姿勢に追従した映像が提示されるため，利用者にあたかも遠隔地に存在し，環境を見回しているかのような感覚を与えることができる．同様の映像提示方式は，アクティブカメラを用いたテレプレゼンスシステム [CB61] や VR 空間での映像提示システム [Sut65] として古くから研究が成されていたが，先に述べた時間遅延や遠隔への情報伝送の問題から，遠隔の実環境を撮影対象とするテレプレゼンスには適用が困難であった．

以上で述べたように，遠隔地の情景を觀賞する実時間ネットワークテレプレゼンスにおいて，全方位映像を用いたシステムはインタラクティブ性や複数の利用者への映像提示に関して優位であると言える．本研究では，これらの優位性を有する全方位映像を用いたテレプレゼンスシステムに焦点を当てる．以下では，観光地や名所の風景をインターネットを介して觀賞可能な実時間ネットワークテレプレゼンスシステムへの応用を想定し，従来の全方位映像を用いたインタラクティブな映像提示によるテレプレゼンスシステムの研究を概観し，それらのシステムの実時間ネットワークテレプレゼンスへの適用可能性について述べる．

## 1.2 インタラクティブな映像提示によるテレプレゼンスの従来研究

観光地や名所の風景をインターネットを介してインタラクティブに観賞するようなアプリケーションを想定した場合に、実時間ネットワークテレプレゼンスシステムに要求される機能について考察する。

まず前節で述べたように、利用者の動作に応じてインタラクティブに遠隔地の映像を提示することで、遠隔地で実際に風景を観賞している時に与えられる視覚情報の再現が可能となり、高臨場感のテレプレゼンスに繋がる。風景の観賞といった用途を想定した場合には、この機能は特に重要であると考えられる。また、そのような利用者の動作に追従した映像提示には、アクティブカメラを用いたテレプレゼンスシステムのように必要とされる映像の取得に時間遅延が伴う構成ではなく、常に利用者に提示する可能性のある視界範囲を撮影し、利用者の動作から必要とされる領域を遅延を感じさせることなく提示可能な構成であるべきである。よって、利用者の動作に応じたインタラクティブな映像提示を行うシステムは、環境の広視野を一度に撮影可能な全方位カメラを用いる必要がある。そして、撮影する風景には静的な物体だけでなく、動的な物体も存在することが想定されるため、動的環境においても動作可能なシステムであることが望まれる。

遠隔地で撮影された映像は、インターネット等のネットワークを介して実時間で利用者に伝送する必要がある。その際、想定するようなアプリケーションでは、映像を送信するシステムと利用者数が常に一対一となることは考え難いため、複数の利用者が同時に遠隔の映像を観賞可能な構成であることが望ましい。

上で述べた実時間ネットワークテレプレゼンスシステムに必要とされる機能をまとめると、以下のようになる。

- A. 全方位の動的環境の仮想化：これにより利用者の動作に応じたインタラクティブな映像提示を可能にする。本論文では、現実環境を撮影し計算機により再現可能な状態にすることから、『仮想化』という用語を用いる。
- B. 映像の実時間ネットワーク伝送：これにより遠隔地で撮影された映像を利用者が観賞可能にする。
- C. 利用者の動作に応じたインタラクティブな映像提示：これにより利用者に遠

隔地に居るような感覚を再現する。

D. 利用者数に対するスケーラビリティ：これにより多数の利用者が同時にシステムを利用し，遠隔を観賞可能にする。

これらの機能を満たすことで，想定するアプリケーションに対して有効な実時間ネットワークテレプレゼンスシステムが実現される。また，高解像度映像の取得や提示，広視野映像提示等も重要な機能と考えられるが，これらの機能は撮像素子の進歩や計算機の高性能化，ネットワークの高速化，表示機器の進歩によって実現されるため，本研究では取り扱わない。

以下では上記の要求機能を踏まえ，全方位動画像を用いたインタラクティブなテレプレゼンスシステムの従来研究を概観する。そして，利用者の動作に対するインタラクティブ性の向上を実現する自由視点画像生成技術について述べ，それらの実時間ネットワークテレプレゼンスへの適用可能性を考察する。

### 1.2.1 全方位映像によるインタラクティブテレプレゼンス

本節では，上記の実時間ネットワークテレプレゼンスに要求される機能を踏まえ，全方位映像を用いたインタラクティブなテレプレゼンスシステムの従来研究を概観する。

現在存在する全方位映像を用いたシステムは，全方位画像を用いてインタラクティブに視線方向の変更を可能とした QuickTimeVR[Che95] を基礎としている。QuickTimeVR はカメラを回転させることで得られる複数の画像を繋ぎ合わせることで全方位画像を生成し，その画像を仮想環境中のカメラ位置を中心とした円筒面にマッピングする。利用者はその円筒面内で自由に視線方向を変更することで，その視線方向に対応した情景の観賞が可能となる。このように，環境を撮影した画像をカメラ位置を中心とした円筒面・立方体・球面にマッピングし，視界画像に入射される光線をその環境マップから取得するという方式は，多くのシステムに用いられている。また，QuickTimeVR は Web ブラウザ上で動作する等，インターネットを介して誰もが簡単に利用できるシステムとなっているが，全方位画像の生成にカメラを回転させて画像を繋ぎ合わせる処理を必要とするため，全方位

の動的環境の撮影は困難であった。

Onoe らのシステム [OYTY98] は QuickTimeVR の課題であった全方位映像の取得を全方位カメラを用いて解決し、利用者の視線方向に追従した映像提示を行うための高速な画像変換方式を提案した。このシステムは双曲面ミラーを用いた全方位カメラ [山澤 96] によって全方位動画像を取得し、HMD と電磁気センサを組み合わせることで利用者の視線方向の変化に合わせた映像提示を行う臨場感の高いシステムである。しかし、全方位カメラと映像提示用 PC は直接接続または、近距離無線通信により接続されておりネットワークを介した実装には到っていなかった。そのため、遠隔地の情景の提示や複数の利用者が同時にシステムを使用することは困難であった。

森田らのシステム [森田 02] は、Onoe らのシステムを基に、ネットワークを介した全方位映像の伝送機構を備えたテレプレゼンスシステムである。実時間での全方位映像のネットワーク伝送により、遠隔監視システムへの応用も可能であることを実証している。このシステムでは全方位カメラで取得した全方位動画像を DVTS ( Digital Video Transport System ) を用いてネットワーク伝送する。Onoe らのシステムと同様に双曲面ミラーを用いた全方位カメラを使用したシステムであるが、カメラ部分に DV カメラを使用することで、高い処理コストを要求される映像の圧縮符号化をハードウェア的に実行する構成となっている。ただし、Peer to Peer 接続によって映像伝送を行うため、複数の PC が同時に映像を受信するのは困難である。そこでこのシステムは 1 台の PC で映像を受信し、その PC で複数人に対しそれぞれの利用者の視線方向の映像を生成している ( 図 5 参照 )。画像の提示には Onoe らのシステム同様、HMD と電磁気センサを組み合わせたものを使用している。

魚眼レンズやミラーによる全方位カメラでは、通常のカメラの画角より大幅に広い画角を撮影することになるため、全方位画像の各画素の角度分解能が低くなることが問題とされてきた。これに対し、通常のカメラとほぼ同等の角度分解能を実現する全方位撮像系として、全方位型マルチカメラシステム [池田 03] がある。この全方位型マルチカメラシステムは、複数のカメラが放射状外向きに配置されており、それぞれが同時に撮影を行うことで、全方位の環境が撮影可能とな

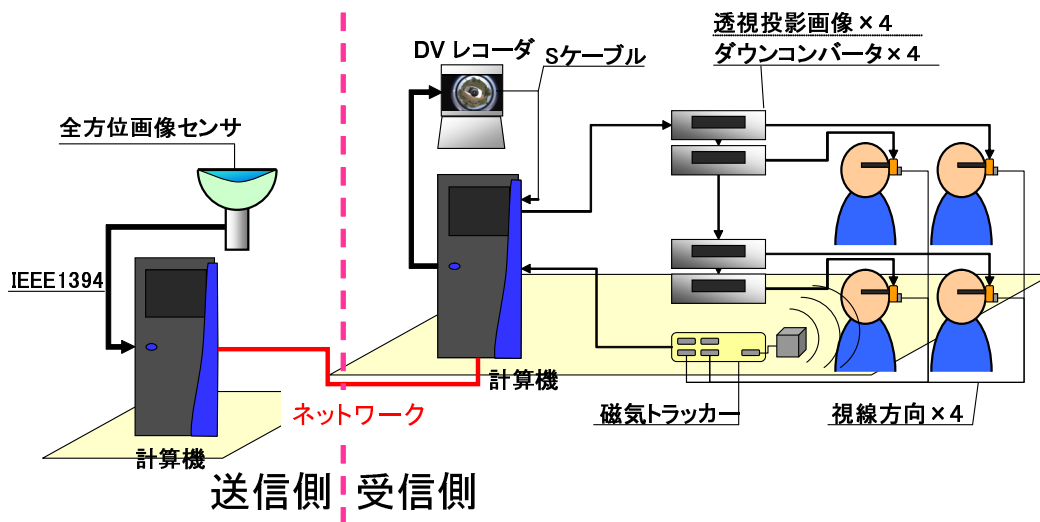


図 5 森田らのシステム構成 [森田 02]

る．池田らはこの全方位型マルチカメラシステムにより高解像度の全方位動画像を取得し，その画像を利用者の視野を覆う球面ディスプレイに投影するシステム [池田 03] を提案している．このシステムでは，予め全方位型マルチカメラシステムを光学的・幾何学的にキャリブレーションしておき，その情報を基にオフライン処理で高解像度な全方位映像を生成・蓄積する．システムの利用者は，球面ディスプレイに投影される映像の視線方向をゲームパッドによって自由に変更することができる．図 6 にシステムの外観を示す．高解像度な全方位画像を使用するため，投影面の広いディスプレイに表示しても画像の粗さを感じ難いというメリットがあるが，環境撮影中の映像をディスクアレイに保持し，撮影後にオフライン処理でモザイクキングを行う必要があるため，予め撮影しておいた映像を再生するような用途に使用が限定されている．

全方位型マルチカメラシステムから得られた全方位動画像を用いた実時間ネットワークテレプレゼンスシステムに PanoVi [YD01] や LiveSPHERE [ビュ03] がある．PanoVi のシステム構成を図 7 に示す．PanoVi は広角な 4 台のカメラを 90 度毎に配置した全方位型マルチカメラシステムにより全周を撮影する．そして撮影された 4 つの画像を一つの画像にまとめ，圧縮符号化後ネットワークを通して遠隔地へ伝送する．伝送時はまだパノラマ画像にはなっておらず 4 つの画像が単純に



図 6 池田らのテレプレゼンスシステムの外観 [池田 03]

並べてあるだけである．伝送された動画像をクライアント PC の GPU ( Graphics Processing Unit ) 機能を用いて立方体にテクスチャマッピングすることによりモザイクングを行う．そして，テクスチャマッピングされた立方体内から任意の方向を見ることによって全周を見回すことができる．映像送信側で複数の全方位映像をモザイクングせず，処理を高圧縮率の映像符号化に集中させることで，高解像度全方位映像の実時間ネットワーク伝送を可能としている．

LiveSPHERE は，池田らと同様の全方位型マルチカメラシステムを用い，PanoVi と同様の GPU 機能を利用した全方位映像のモザイクング処理を行う．ただし，ネットワークで伝送するのは，システムの利用者から伝送された視線情報を基に生成した視界画像である．利用者の必要としている部分の画像のみを伝送することで，高解像度全方位映像全体の伝送を回避している．しかし，視線情報の変更により映像送信側と利用者間の距離に依存した時間遅延が生じることや，利用者数が増加することによるネットワーク負荷や視界画像生成処理の負荷が増大することが問題となる．

PanoVi と LiveSPHERE は全方位動画像の取得からネットワーク伝送・提示まで

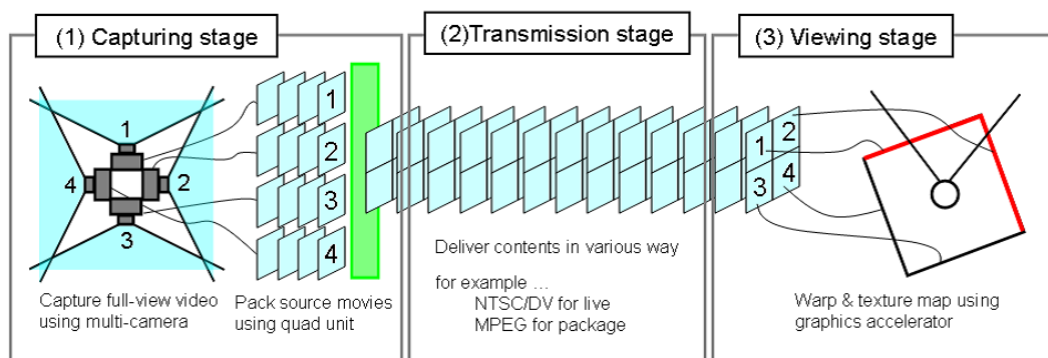


図 7 PanoVi のシステム構成 [YD01]

を実時間で行うことができるシステムであるが，利用者が増加すると伝送に必要なネットワーク帯域や処理が増加するといった点において解決すべき課題が残る．

以上に挙げた全方位画像及び映像を用いたインタラクティブな映像提示によるテレプレゼンスシステムは，共通して利用者数に対するスケーラビリティに関して問題を有する．ネットワークを用いたテレプレゼンスシステムでは，映像送信側は利用者それぞれに対して独立な処理を行うのではなく，単一の処理により複数の利用者に映像を伝送できる，利用者数に依存しない機構を有するべきである．

これら全方位映像を用いたテレプレゼンスシステムでは，利用者は視線方向を自由に変更可能となり，HMD と頭部の姿勢を計測するセンサを用いた場合には視線に追従した映像観賞が可能となる．しかし，利用者の動作に応じた映像提示を行うためには，視線方向だけでなく視点位置にも対応した映像提示が必要である．

次節では，視点位置を自由に変更し遠隔の映像を観賞可能なテレプレゼンスシステムの従来研究および，コンピュータビジョンによる自由視点画像生成技術とそのテレプレゼンスシステムへの実装例について述べる．

### 1.2.2 自由視点画像生成によるテレプレゼンス

遠隔地の情景をその場にいるような感覚で観賞可能とするためには，利用者の頭部の動きに伴う視線方向・視点位置に対応した映像を提示することが望まれる．

インタラクティブに視点位置を変更するための方法として，予め視点位置が移

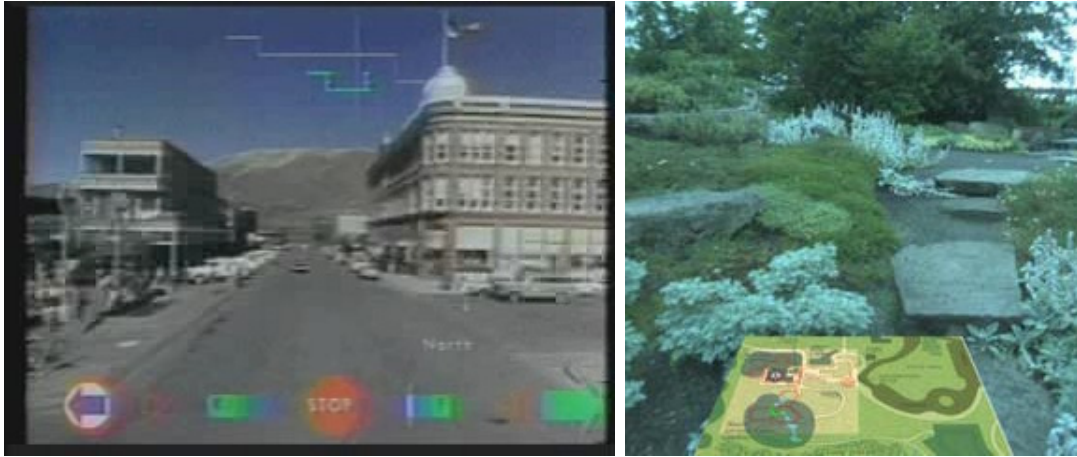


図 8 視点位置に応じた画像の選択的提示によるシステム（左：Aspen MovieMap [Lip80]，右：Uyttendaele らのシステム [UCK+04]）

動するあらゆる場所において映像を記録しておき，利用者の視点位置に応じて選択的に画像を提示する方法が古くから用いられている．その代表的な研究は，Aspen MovieMap[Lip80]である．このシステムでは全方位カメラを車に搭載し，市街地の様々な経路を走行しながら映像を光ディスクに記録する．そして，利用者は車に乗っているような感覚で経路の選択や視線方向の変更ができる（図 8 左参照）．同様のシステムには，Uyttendaele ら [UCK+04] や Taylor[Tay02] のシステムがあり，全方位カメラにより撮影した全方位映像を基に，HDR(High-Dynamic Range) 画像生成や映像の安定化，Structure-from-Motion による経路の自動生成等，提示映像の高画質化や有用性の向上のための研究が成されている．これらのシステムは，映像を提示する必要のある利用者の視点位置が予め分かっている場合には有効であるが，視点の切り替わりに不連続が生じることや，提示映像の時間的な連続性が考慮されていないため環境中の動物体が正しく表現できないことが問題とされている．また，映像を取得する必要のある範囲が分かっていたとしても，その範囲が広い場合には画像のデータ量や撮影のためのコストが膨大となる．

これらの問題に対して，近年，現実環境を複数の視点で撮影した画像群から，それらの視点位置とは異なる位置における視界画像を生成する自由視点画像生成技術が盛んに研究されている．自由視点画像生成技術は，任意の視点位置の



視界画像を生成するため，利用者の視点位置の変化に対して連続的な映像提示が可能であることや，選択的に画像を切り替える方法に比べ密な環境の撮影が不要といった利点がある．この自由視点画像生成手法は，画像生成の際に環境の幾何学的情報を明示的に算出するか否かにより，3つの手法に分類できる．環境を撮影した多視点画像からその環境の3次元形状を復元することにより自由な視点の視界画像を生成するモデルベースレンダリング（MBR：Model-Based Rendering）[SBK03, KNR97, MT02, SD98, ZKU<sup>+</sup>04, 高橋06]，環境の3次元形状を明示的に復元することなく画像を生成するイメージベースレンダリング（IBR:Image-Based Rendering）[CW96, SD96, 遠藤02, YIS04, LH96, GGSC96, 山口01, 岡03, MBR<sup>+</sup>00, IHA03, FWZ05, 稲本04]，モデルベースレンダリングとイメージベースレンダリングを組み合わせたハイブリッドレンダリング [KKO03, KO03, YSK<sup>+</sup>02, STYY00, TYY02] の3つである．以下では，これら3つの手法について，その特徴や従来研究について概観する．

#### [モデルベースレンダリング]

モデルベースレンダリングでは，自由視点画像の生成のために環境の3次元形状を復元する．環境の形状を用いるため，視点移動に関する自由度や視点移動に伴う物体の隠蔽関係の表現に優れる．また仮想化された環境にさらに仮想物体を配置することや，仮想化された物体の移動が容易に行える．しかし，モデルベースレンダリングの多くはマルチベースラインステレオ法 [OK93] や視体積交差法 [Bak77] により環境の形状を推定するため，自然景観のように複雑な形状を持つ環境や物体はその復元が困難とされる．また，一般的に環境の3次元形状を復元するには膨大な計算コストが掛かる．

環境の3次元形状を取得するために，レーザーレンジファインダ（図9参照）を使用する方法 [浅井05] も提案されているが，物体の形状を高精度に取得可能である反面，形状の計測に時間がかかるため動的環境を実時間で仮想化するような用途には不向きである．

コンピュータビジョンの技術を用いて多視点画像から動的環境を仮想化する研究の代表例に，Kanadeらによる Virtualized Reality プロジェクト [KNR97, SBK03]

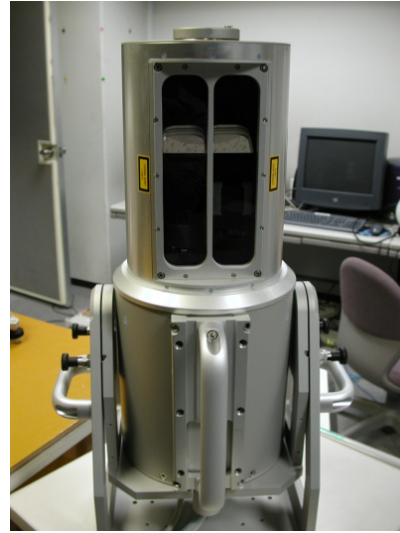


図 9 レーザレンジファインダの例 (左: Leica 社製 TCR1105XR, 右: Riegl 社製 LMS-Z360)

がある。これらの研究では、位置・姿勢が既知のカメラを複数配置した 3D Room と呼ばれる室内環境 (図 10 参照) において、それらのカメラで同時に撮影された動画像を用いて室内の動物体の 3 次元復元を行った。Kanade ら [KNR97] はマルチベースラインステレオ法により各画像の奥行き値を計算し、それらを 3 次元モデルとして統合した。Saito ら [SBK03] はステレオ法による誤差をシルエット制約を加えることにより防止し、物体のテクスチャを Morphing により合成することで視点の移動に伴うテクスチャの切り替わりの違和感を低減した。動的環境の仮想化後、あらゆる視点位置からその環境を観賞することが可能となる。

同様のカメラ配置では、視体積交差法により対象物体の近似的な 3 次元形状を得る手法が用いられることが多い。視体積交差法は撮影空間をボクセルと呼ばれる立方体の集合として表現し、そのボクセル空間内で多地点でのカメラの光学中心とシルエット画像から成る視体積の積集合空間 (Visual Hull) を計算する。Visual Hull は、対象物体を様々な方向から撮影したシルエット画像の量を増やすことによって対象物体の真の 3 次元形状に近づく性質がある。ステレオ法に基づく手法と異なり対応点探索の問題が無く、ボクセル空間の解像度や使用する画像数にも依るが、処理速度はステレオ法による 3 次元形状復元と比較すると高速で

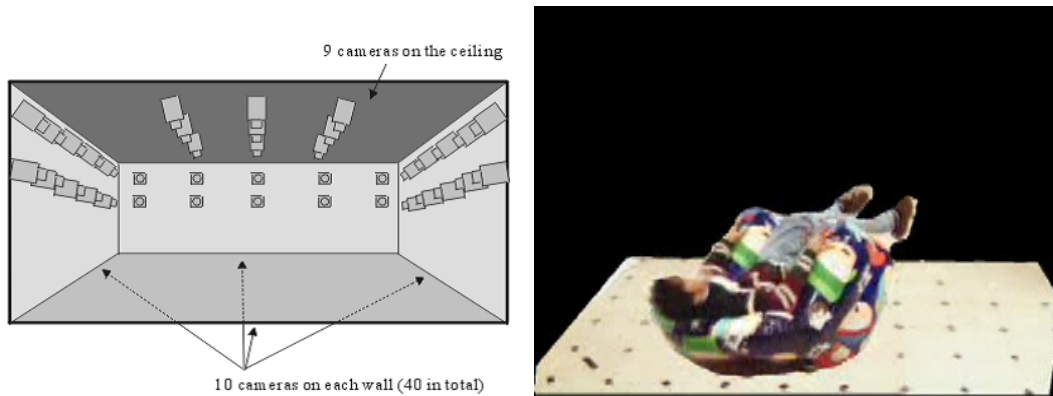


図 10 3D Room と 3 次元形状復元した動物体 [SBK03]

ある．ただし，対象物体に凹な部分が存在する場合には，その部分は正しい 3 次元形状が得られないという制約がある．Matsuyama らは [MT02]，各カメラのシルエット画像を視体積の内外判定が高速に行われるよう射影変換し，PC クラスタによって実時間での撮影対象の 3 次元復元を行った．Seitz ら [SD98] は，ボクセル空間内の各ボクセルを各カメラに投影した時にシルエットの内外を判定するのではなく色の一致 (Photo-consistency) を評価することで，3 次元復元を行った．この Photo-consistency により得られた形状は，Visual Hull に対して Photo Hull と呼ばれる．これら視体積交差法に類する手法は Kanade ら [KNR97] や Saito ら [SBK03] の手法同様に，対象物体の詳細な 3 次元形状を得ることを目的としており，特定の領域を集中して撮影する必要があるため，環境の仮想化が可能な範囲が限定されやすい．遠隔の情景を観賞するような目的でのこれらの手法の適用は，広範囲の撮影空間に対応するための計算量やデータ量の多さから非常に困難と言える．

遠隔の情景をインタラクティブに観賞するテレプレゼンスの用途に適した手法としては，Zitnick ら [ZKU<sup>+</sup>04] や高橋ら [高橋 06] の研究がある．Zitnick らはカメラアレイにより蓄積した多視点映像を基に，セグメンテーションベースドステレオによる奥行き推定を行い，各カメラの奥行き画像を蓄積する．映像提示の際は，自由視点位置に隣接した二つのカメラの奥行き画像から二つの 2.5 次元モデルを生成し，それらを自由視点画像面に投影することで自由視点画像を生成する．



図 11 セグメンテーションベースドステレオにより推定された奥行き（左）と自由視点画像例（右） [ZKU+04]

環境の平面集合仮定によりテクスチャの少ない環境においても安定した奥行き推定を行い、自由視点画像のレンダリングの際には、アーティファクトを防止するために奥行き値が大きく変動する領域に平滑化処理を施すことで高品質な自由視点映像を観賞可能としている（図 11 参照）。しかし、セグメンテーションベースドステレオの計算コストの高さから映像の取得から提示までを実時間で行うには至っていない。高橋らの手法は、自由視点画像面に正対する焦点面（Focal plane）を設定し、その焦点面を奥行き方向に変化させ、各奥行きにおける焦点面上の色の一致度を測ることで環境の奥行きを推定する。焦点面には多視点画像を投影するが、その処理を近年高性能化が進む GPU を用いることで高速に行い、映像の撮影から提示までを実時間で行う。実時間ネットワークテレプレゼンスへの適用可能性が高い手法ではあるが、安定した奥行き推定にはカメラ間の距離を十分に短くする必要があるため、自由視点の移動可能な範囲はごく狭い領域に限られる。

#### [イメージベースドレンダリング]

イメージベースドレンダリングは、環境の 3 次元形状を明示的に復元することなく、撮影画像の変形・合成により自由視点画像生成を行う。環境の 3 次元形状を持たないため、自由視点画像を生成するための処理は環境の複雑さに依存せず、多くの手法ではモデルベースドレンダリング手法に比べ高速にレンダリングが可能である。しかし、環境の形状に関する情報を持たないため、物体の隠蔽関係の

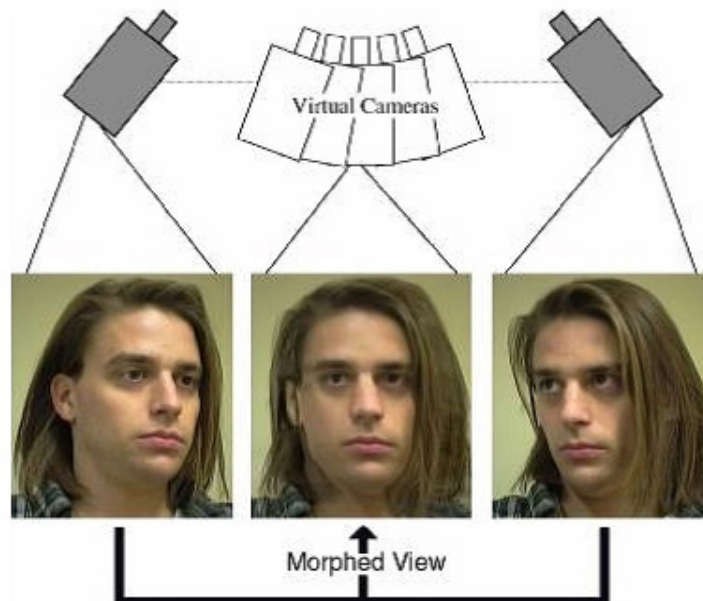


図 12 View Morphing [SD96]

表現は不得意であり，そのような領域では画像にぼけやアーティファクトが発生することが多い．

イメージベースレンダリングの代表的な手法は，Warping や Morphing によって自由視点画像のレンダリングをする手法 [CW96, SD96, 遠藤 02, YIS04] である．これらの手法の基本的な戦略は，2 枚以上の入力画像間の対応点を与え，中間視点位置において対応点がどの位置に現れるかを計算し，さらにその中間視点における対応点の色を入力画像の色をブレンドすることによって決めるというものである（図 12 参照）．これらの手法は対応点を与えられているなら高速に自由視点画像生成が可能であるが，モデルベースレンダリング同様，対応点を自動的にかつ正確に与えることは困難である．Yamazaki らの手法 [YIS04] は，入力画像に関する知識なしに自動的に対応点を与えることを可能にしている．しかし，対応の計算に多くの計算コストが必要となる．

対応点を必要としないイメージベースレンダリングに，空間中を伝播する光線を記録し，自由視点位置において視界画像面を通過する光線をリサンプリング

することで自由視点画像生成を行う手法がある [LH96, GGSC96, 山口 01, 岡 03] . Levoy ら [LH96] や Gortler ら [GGSC96] は , Light field rendering と Lumigraph という名の手法をほぼ同時期に発表し , 空間中を伝播する光線は 4 次元パラメータで表現可能であることを示した . 山口ら [山口 01] は , Light field rendering の考えを応用し , 全方位カメラによる効率的な光線情報の取得及び , 視線方向も変更可能な自由視点画像生成手法を提案した . 岡ら [岡 03] は , EPI による光線情報の補間法とベクトル量子化による効率的な光線情報の圧縮法を提案した . これらの手法は , 他の手法に比べ環境を密に撮影する必要があることや , 撮影時のカメラの位置・姿勢情報が必要であること等の制約はあるが , 撮影環境に関する情報は一切必要なく , 自由視点画像生成も高速に行うことが可能である .

自由視点画像面を通過する光線をリサンプリングする Light field rendering に対し , 自由視点からその画像面を通過する光線を設定し , その光線の色を撮影画像から計算する Ray-casting による手法が提案されている . Matusik ら [MBR<sup>+</sup>00] は入力シルエット画像とエピポーラ拘束を基に , Visual Hull が自由視点画像上にどのように投影されるかを計算する手法を提案した ( 図 13 参照 . ただし , 図右の人物以外の周辺環境は予め作成された 3DCG である ) . この手法により計算される自由視点画像は , 視体積交差法によるそれとほぼ同等であるが , ボクセルを用いないためボクセル空間の空間分解能を決定する必要がないという利点がある . また , Matusik らはこの手法を用いて実時間自由視点画像生成が可能なシステムを構築しており , 高速な処理が可能であることを示している .

同様のアプローチにより , シルエット制約に依らず自由視点画像を生成する手法が Irani ら [IHA03] や Fitzgibbon ら [FWZ05] によって提案されている . Irani らは , 自由視点からの光線を設定し , その光線上の点を多視点画像に投影した時 , 最も色が一致する点を探索する . そして , その最も色が一致した点の色をその光線の色として , 自由視点画像にレンダリングする . この処理は , ステレオ法によって奥行き値を評価する処理とほぼ同等と言える . Fitzgibbon らは Irani らの手法を改良し , 入力多視点画像上に存在しないようなテクスチャパターンが自由視点画像上に発生することを防止するエネルギー関数を定義した . 明示的に環境の 3 次元形状を復元するわけではないが , 本質的にはステレオ法と同等であり , その

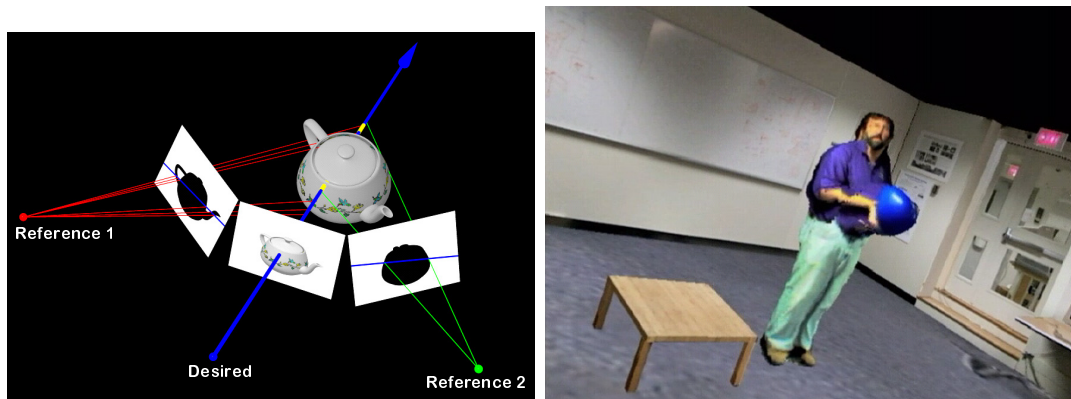


図 13 Image-Based Visual Hull [MBR+00]

計算コストは高い。

稲本ら [稲本 04] は、複数のイメージベースレンダリング手法を組み合わせることによりサッカーシーンに適した自由視点画像生成手法およびその手法を用いたシステムを提案している。この手法の概要を図 14 に示す。まず、撮影された画像を背景差分処理により動的領域と静的領域に分離する。動的領域としてサッカープレイヤーやボール、静的領域としてサッカーフィールドやゴールポスト、背景が得られる。サッカーシーンにおいて、サッカーフィールドやゴールポスト、遠景は一つの平面又は複数の平面の集合と見なす事ができる。そのサッカーフィールドやゴールポストの平面に対しては Homography を用いて自由視点位置における画像を生成する。遠景に関してはモザイクと自由視点から見えている領域の切り出し処理により自由視点画像とする。サッカープレイヤーやボールはエピソード拘束を用いた対応付けを行い Morphing によって自由視点画像生成を行う。最終的にそれら生成された自由視点画像を統合することにより提示画像とする (図 15 参照)。サッカーシーンに特化することで高品質な自由視点画像を生成することが可能であることを示している。

#### [ハイブリッドレンダリング]

モデルベースレンダリングとイメージベースレンダリングとを組み合わせたハイブリッドレンダリングによる手法も様々提案されている。Koyama ら [KKO03]

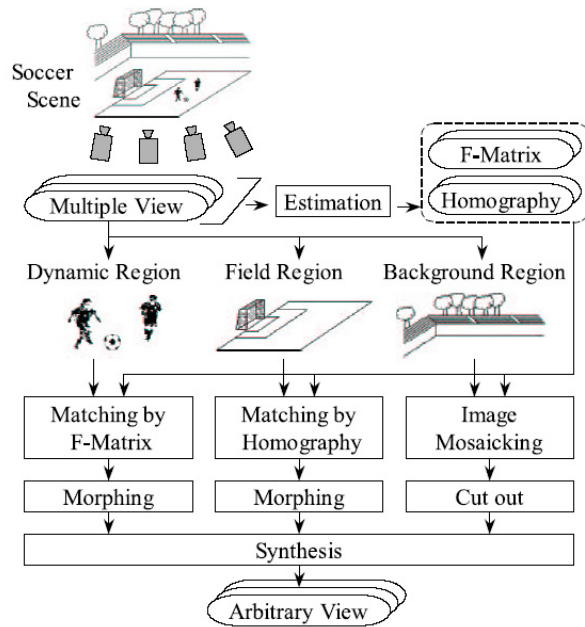


図 14 稲本らの自由視点画像生成手法の概要 [稲本 04]



図 15 稲本らの手法により生成された自由視点画像 [稲本 04]





図 16 ビルボーディング技術を用いた自由視点画像生成 [KKO03]

は、サッカーシーンを自由視点で観賞することを目的とし、ビルボーディング技術を利用した自由視点画像生成手法や、その手法を利用してサッカーのライブ映像を自由な視点で観賞可能なシステムを提案している。サッカーシーンにおいて、観客席からサッカープレイヤーまでの距離はある程度離れており、サッカープレイヤーの形状を正確に得る必要はないと考えられる。そこでこの手法は、推定したプレイヤーの位置に平面のオブジェクトを描き、入力画像から得られるサッカープレイヤーの画像をテクスチャマッピングし、視点位置に応じてその平面の向きを変化させる簡易なプレイヤーモデルによる自由視点画像生成を行った。プレイヤー以外の物体に関しては、予め3DCGで作成したものを配置している。また、この手法を利用したライブ映像配信が可能なネットワークテレプレゼンスシステム実現しており、インタラクティブにサッカー観賞が可能であることを実証している(図16参照)。しかし、このシステムは利用者の視点・視線情報をサーバに送信し、その情報を基にサーバは利用者に必要なデータを送信する方式をとっているため、利用者数に対するスケーラビリティに欠ける。

Koyamaら [KKO03]と同様に、Kitaharaら [KO03]は平面の集合によって近似的に対象物体の3次元形状を表現することで自由視点画像を生成する手法を提案

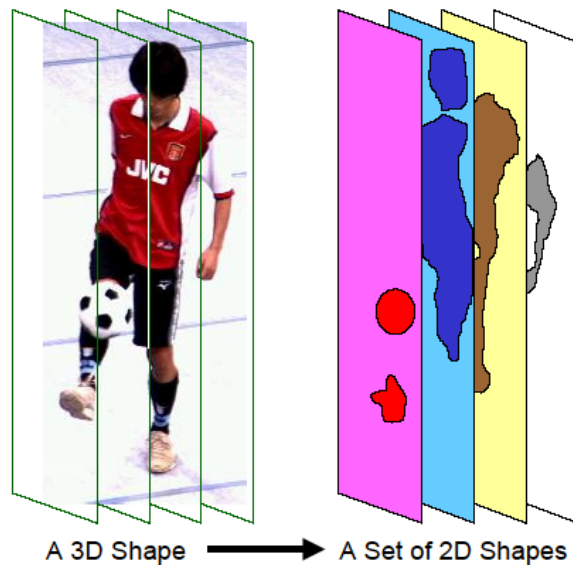


図 17 平面の集合によって近似的に三次元形状を復元する手法 [KO03]

している（図 17 参照）。ボクセル空間等により物体を表現するより計算コストが少なく高速に処理が可能である。また，大規模空間で自由視点画像を生成する際に，対象物体と視点位置との位置関係により近似的な 3 次元形状の空間解像度を变化させ 3 次元形状復元の冗長性を減少させている。サッカー場のような上方からの映像が取得できる環境でなければ，対象物体の位置が安定して正しく推定できないため，多視点画像を射影変換するコストが増加することが考えられる。

Shimamura らの手法 [STYY00] は全周ステレオパノラマ画像を用いてカメラから物体までの奥行きを求め，2.5 次元モデルによって全方位の環境を表現する（図 18 参照）。この手法で用いられているカメラは全周ステレオパノラマ動画を撮影可能であり動的環境にも対応できる。しかし，この手法では 1 つの視点位置から得られた全周ステレオパノラマ画像を用いるため，生成する画像の視点位置が撮影位置から離れるにつれて画像に歪みが生じる。

冨手らの手法 [冨手 02] は多地点で全方位画像を撮影し，それらの画像に対応点を与えることで環境の疎な 3 次元形状を復元し，その形状に対するテクスチャを Morphing することによって生成する（図 19 参照）。対応が既知な点の三次元位置の計算と Morphing によって自由視点全方位画像を生成可能するため，ビデオ

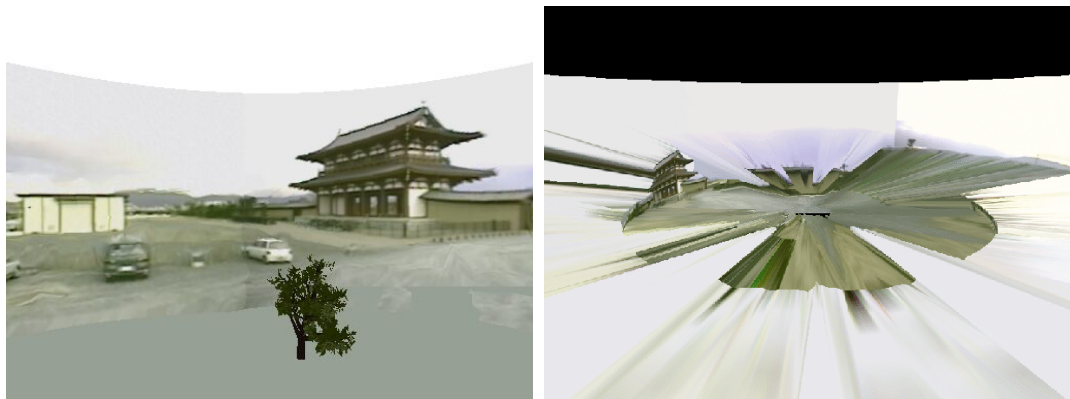


図 18 環境の 2.5 次元モデルと自由視点画像 [STYY00]

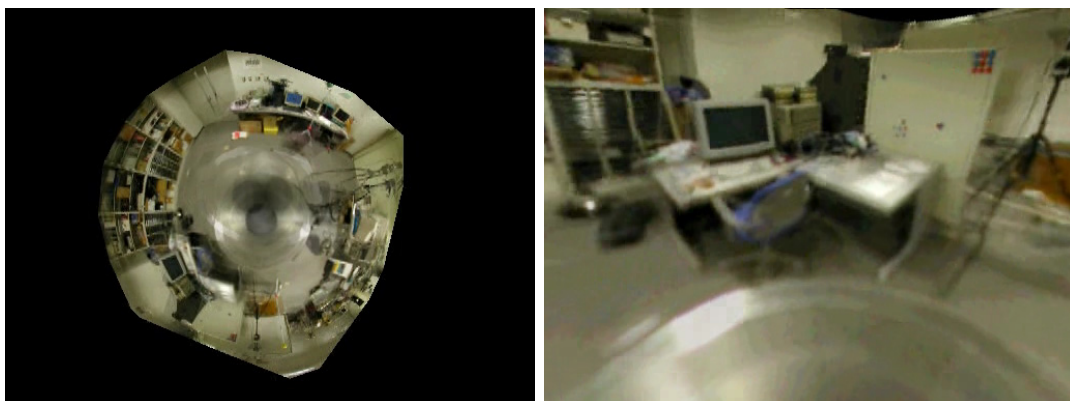


図 19 全方位自由視点画像と平面透視投影画像 [富手 02]

レートでのレンダリングが可能である．しかし，予め対応点を与える必要があるため，動的環境に適用するのは困難である．

以上のように，自由視点画像生成手法は多種存在し，その特徴も様々である．次節では，実時間ネットワークテレプレゼンスシステムへの自由視点画像生成手法の適用可能性を考察し，本研究の位置づけ及び研究方針について述べる．

### 1.3 本研究の位置付けと方針

前節では、実時間ネットワークテレプレゼンスシステムに必要な機能を以下のように定義し、これらの機能を踏まえた上で、全方位映像を用いたテレプレゼンスシステムの従来研究及び、よりインタラクティブな映像提示を可能にする自由視点画像生成技術の従来研究を概観した。

- A. 全方位の動的環境の仮想化
- B. 映像の実時間ネットワーク伝送
- C. 利用者の動作に応じたインタラクティブな映像提示
- D. 利用者数に対するスケーラビリティ

本節ではこれらの機能を持つテレプレゼンスシステムの実現のための方針を従来研究から得られる知見を基に述べ、本研究の位置づけを述べる。

まず、全方位の現実環境を撮影するにはこれまで述べてきた通り全方位カメラ [COH85, YNY91, 山澤 96, KYI+98, 棚橋 02, 池田 03] を使用する方法が有効であると考えられる。全方位カメラを使用することにより、視線方向の自由度を持つため、インタラクティブな画像提示が可能になり、複数の利用者に対する画像提示も行うことができる。映像のネットワーク伝送の方法や全方位カメラの種類による実時間映像処理の適性に議論の余地があるが、全方位カメラを使用することによるメリットは大きい。従来提案されているシステムでは、要求機能 A,B そして C の視線方向に関して対応している森田らのシステム [森田 02] が最も要求機能を多く満たすシステムと言える。PanoVi[YD01] は、ほぼ森田らのシステムと同等の機能を有しているが、森田らのシステムは利用者の動作に応じた映像提示を行うためのインタフェースとしてHMDと姿勢センサを用いている点で優れている。LiveSPHERE[ビュ03] は、映像送信側で利用者それぞれに対する処理を行うため、利用者数に対するスケーラビリティに欠けることに加え、利用者数が増加すると負荷の増大によりインタラクティブ性が低下することも考えられる。また、Koyama らのシステム [KKO03] は、サッカーシーンにおいては A~C の機能を満たしていると言えるが、本研究で想定する遠隔の風景を観賞するようなアプ

リケーションに対しては、予め周辺環境を3DCGで作成するようなことが困難であるため、動物体に関する映像提示のみに限定される。以上から、森田らのシステムを基礎として、利用者の視点位置に応じた映像提示及び、利用者数に対するスケーラビリティを確保することができれば要求機能を全て満たすシステムが実現できると考えられる。

利用者の視点位置に応じた映像提示には自由視点画像生成技術を用いる必要がある。しかし、自由視点画像生成技術を導入することによりその他の機能が制限されることがないようにしなければならない。前節において自由視点画像生成技術を大きく3つに分類し、それらの手法について述べた。ここでは、さらにそれぞれの手法が実時間ネットワークテレプレゼンスシステムに適用可能かどうかを考察する。

まず、モデルベースドレンダリングについて考える。レーザーレンジファインダを用いる手法[浅井05]は三次元形状の計測に時間がかかるため動的環境に適用できない。コンピュータビジョンによる手法のうち、ステレオ法による距離推定を行う手法[KNR97, SBK03, ZKU<sup>+</sup>04]は計算コストや対応点探索の問題から実時間処理は難しい。ボクセル空間において Visual Hull や Photo Hull を評価する手法[MT02, SD98]はボクセル数を少なくすることで実時間処理を可能にできるが、想定するアプリケーションでは撮影する空間は広範囲のためボクセル数は膨大となり、実時間処理は困難となる。高橋らの手法は[高橋06]、カメラの相互投影の問題から全方位カメラを用いたシステムには適していない。よって、モデルベースドレンダリングは実時間ネットワークテレプレゼンスシステムには適用困難である。

次に、イメージベースドレンダリングについて考える。Warping や Morphing による手法は[CW96, SD96, 遠藤02]処理は高速に行えるが対応点が必要であり動的環境への適用が難しい。対応点を自動で与える手法[YIS04]は計算コストがかかり実時間での処理は難しい。空間を伝播する光線を記録する手法[LH96, GGSC96, 山口01, 岡03]は密に画像を取得する必要がある、同時刻の多視点画像を必要とする動的環境の対応は困難である。Matusik らの手法[MBR<sup>+</sup>00]は、シルエット画像を基に Visual Hull の自由視点画像上への投影を計算するため、Visual Hull

以外の環境については対応できない。Irani ら [IHA03] や Fitzgibbon ら [FWZ05] の手法は、本質的にステレオ法と同等であり、計算コストの問題から実時間処理は困難である。稲本らの手法 [稲本 04] はサッカーシーンに特化されており他の環境に適用するのは難しい。よって、イメージベースドレンダリングも実時間ネットワークテレプレゼンスシステムに適用困難である。

次に、ハイブリッドレンダリングについて考える。Koyama らの手法 [KKO03] は視点位置と対象物体が離れていることを仮定しているため、システムの利用環境が限定されることや、動物体以外の環境への対応がなされていない。Kitahara らの手法 [KO03] は、Koyama らの手法同様、動物体以外の環境への対応は考慮されていない。Shimamura らの手法 [STYY00] は、仮想環境の構築に時間がかかり実時間でのテレプレゼンスは難しい。冨手らの手法 [冨手 02] は、環境が静的であることを仮定しているため動的環境に適用困難である。よって、これらの手法も実時間ネットワークテレプレゼンスシステムには適用困難である。以上の考察から、これまで提案されてきた手法を単純に用いるだけでは目標とするテレプレゼンスシステムは実現困難であることが分かる。

そこで、撮影環境中の静的領域と動的領域を分離し、それぞれの領域に適した手法を適用することを考える。また、オフラインで静的領域の自由視点画像生成に必要な処理を行っておき、動的領域のみを実時間で処理する方法も考慮する。これらを考慮すると、撮影環境中の動的領域については Matusik ら [MBR<sup>+</sup>00] の手法を、静的領域については冨手らの手法 [冨手 02] を用いてそれぞれ自由視点画像生成を行い合成することで環境全体の自由視点画像が生成できると考えられる。しかし、これらの手法は広帯域なバス上の映像や予め撮影された静止画像に対して用いられており、実時間ネットワークテレプレゼンスシステムに適用するには、ネットワーク伝送に要するオーバーヘッドや各手法の処理時間を考慮して、より高速な画像生成処理を行う必要がある。

本研究では、遠隔の観光地や名所の風景をインターネットを介してインタラクティブに観賞するアプリケーションへの応用を想定した時に要求される機能 A ~ D を全て満たす実時間ネットワークテレプレゼンスシステムの実現を目的とする。目標のテレプレゼンスシステム実現のため、それぞれの機能に対し以下のア

アプローチをとる．

A. 全方位の動的環境の仮想化

多地点に配置した全方位カメラにより環境全体を同期撮影

B. 映像の実時間ネットワーク伝送

多視点画像の圧縮符号化及び実時間ネットワーク伝送

C. 利用者の動作に応じたインタラクティブな映像提示

多視点全方位映像からの実時間自由視点画像生成

D. 利用者数に対するスケーラビリティ

マルチキャストプロトコルによるネットワークを介した多地点同報通信

本研究ではまず，Onoeらのテレプレゼンス方式 [OYTY98] を基礎とし，マルチキャストプロトコルを用いた映像伝送により，利用者数が増加したとしてもネットワーク負荷やサーバの計算コストの増加しないハイスケーラブルな実時間ネットワークテレプレゼンスシステムを構築する．また，システム単体としても従来の全方位映像を用いたテレプレゼンスシステムよりユーザビリティに優れるシステムとする．そして，そのハイスケーラブルな実時間ネットワークテレプレゼンスシステムを基に，多視点全方位画像からの実時間自由視点画像生成処理を加え，要求機能 A～D を全て満たす実時間ネットワークテレプレゼンスシステムを実現する．

## 1.4 本論文の構成

本論文の構成は以下のとおりである．第2章では，利用者の増加に対応可能なスケーラビリティを有する実時間ネットワークテレプレゼンスシステムを提案する．配信する全方位映像は利用者の視線方向に非依存である特徴と，ネットワーク上で同報通信を可能にするマルチキャストプロトコルを利用することでハイスケーラビリティを実現可能であることを示す．第3章では，第2章で提案したシステムを基に自由に視点位置を変更可能な自由視点テレプレゼンスシステムを提案する．多地点に配置した全方位カメラからの映像を用いて実時間で自由視点画像を生成する手法を提案し，実際にプロトタイプシステムを構築する．そして，実験により生成された自由視点画像の定量的画質評価および提案システムの有効性の確認を行う．最後に第4章で，本研究を総括する．



## 2. 全方位映像のマルチキャストによる実時間ネットワークテレプレゼンス

### 2.1 序言

本章では，マルチキャストプロトコルによって映像を伝送し，多地点で複数人が同時にインタラクティブ観賞可能な実時間ネットワークテレプレゼンスシステムについて述べる．従来の全方位映像を用いたテレプレゼンスシステム [Che95, OYTY98, 森田 02, 池田 03, YD01, ビュ03] は，蓄積または実時間で伝送された全方位映像中から利用者の視線方向に対応する領域を切り出し，変換することで視線方向に対してインタラクティブな映像提示を実現している．遠隔地の全方位映像をネットワークで実時間伝送し，視界画像の生成処理を高速に行うことで，同一地点の複数人でそれぞれが別々の方向を観賞することが可能なシステムも提案されている [森田 02]．しかしこれらのシステムは，遠隔への映像伝送の際にサーバ側で利用者の視界画像を生成・伝送する機構 [ビュ03] や P2P 接続による映像伝送機構 [YD01, 森田 02] を用いているため，多地点の複数人がシステムを同時に使用することは，サーバの処理コストやネットワーク負荷の面から困難であった．また，これらのシステムは一般にはあまり用いられないような研究用の機材による映像提示を行っており，実用的とは言えなかった．

本研究では，映像を配信するサーバ側では利用者数や利用者の視線方向に依存する処理を排除し，ネットワーク上で一対多の同報通信を可能とするマルチキャストプロトコルによる全方位映像の配信を行うことで，従来システムのスケールビリティの問題を解決する．また，システムのユーザビリティについても改善を行い，Web ブラウザによる全方位映像の提示を行うことで，誰でも簡単にテレプレゼンスを体験できるシステムを提案する．

以下では，本研究で提案する実時間ネットワークテレプレゼンスシステムの基礎となるテレプレゼンス方式である Onoe らのシステム [OYTY98] について述べ，マルチキャストプロトコルによる映像伝送方式について概説する．そして，提案システムの詳細な説明の後，実験によりその機能を実証する．

## 2.2 全方位画像の平面透視投影変換によるテレプレゼンス方式

本節では，全方位画像を高速に平面透視投影変換することにより，利用者の視線方向に追従した映像提示を可能とした Onoe らのテレプレゼンス方式 [OYTY98] について概説する．このテレプレゼンス方式は，現存する多くのシステムで同様の方式が用いられており，計算機の高性能化が著しい今日においてもその重要性に変わりはない．本研究においても，Onoe らの方式を基にしたシステムを提案する．以下ではまず，Onoe らのテレプレゼンスシステムで用いられる全方位カメラ HyperOmni Vision について述べる．そして，HyperOmni Vision により得られる全方位画像からのハードウェア機能を利用した平面透視投影画像の実時間生成手法について述べる．

### 2.2.1 全方位視覚センサ HyperOmni Vision

本節では，全方位視覚センサ HyperOmni Vision [山澤 96] について述べる．HyperOmni Vision は，鉛直下向き of 双曲面ミラーと鉛直上向き of カメラから構成され，周囲 360 度の画像を一度に取得できる．図 20 に HyperOmni Vision の外観とその光学系を示す．ミラー部の曲面は回転二葉双曲面を用いる．図 20 に示すように二つの焦点  $O_M, O_C$  を持ち， $O_C$  にカメラのレンズ中心が位置するようにミラーをカメラに取り付ける． $Z$  軸を鉛直軸とする三次元世界座標系  $(X, Y, Z)$  を考えると，双曲面ミラーの面，ミラー焦点  $O_M$ ，およびカメラのレンズ中心  $O_C$  は式 (1) で表せる ( $a, b, c$  は双曲面の形状を定義する定数) ．

$$\begin{aligned} \text{ミラー曲面} & : \frac{X^2 + Y^2}{a^2} - \frac{Z^2}{b^2} = -1 \\ \text{ミラー焦点 } O_M & : (0, 0, +c) \\ \text{レンズ中心 } O_C & : (c = \sqrt{a^2 + b^2}) \end{aligned} \tag{1}$$

また，ミラー内側の焦点  $O_M$  に集まる像は双曲面ミラーに反射して，外側の焦点  $O_C$  に集まる．ここで， $O_C$  にレンズ中心を置いた焦点距離  $f$  のカメラで全方位画像を撮影すると，双曲面の特性により，三次元空間中の点  $P(X, Y, Z)$  と全方位画像上の写像点  $p(x, y)$  の間には式 (2) の関係が成立する．

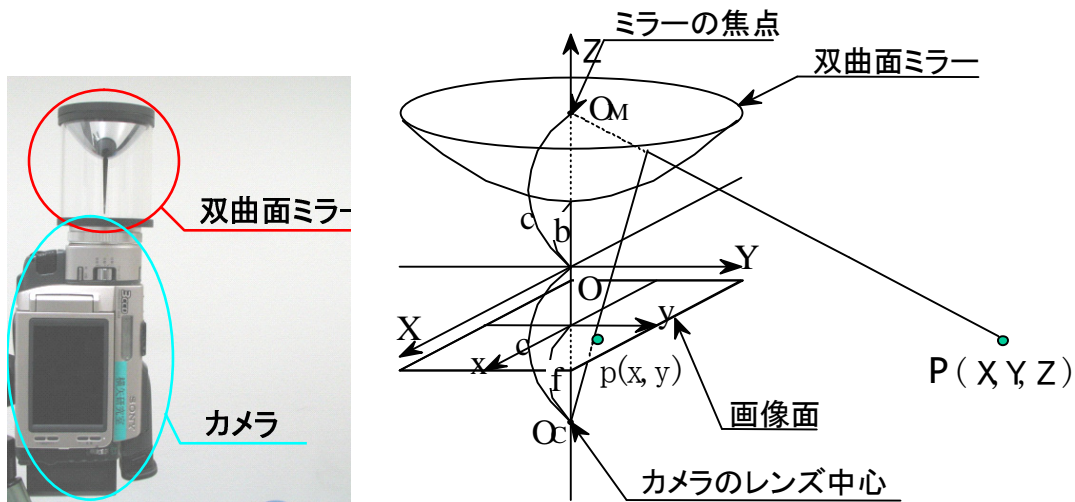


図 20 HyperOmni Vision の概観 (左) と光学系 (右)

$$x = \frac{X f (b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}} \quad (2)$$

$$y = \frac{Y f (b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}}$$

式 (2) より，焦点  $O_C$  で撮影された全方位画像上の点が，焦点  $O_M$  から見て 3 次元空間中のどの方向に対応するかが分かるため，全方位画像を 3 次元空間中の平面に再投影することにより平面透視投影画像を生成可能である．

図 21 に HyperOmni Vision で撮影された全方位画像とその全方位画像の一部を平面透視投影変換した画像を示す．撮影された全方位画像は円状に歪んでいるが，それを平面透視投影変換した画像には歪みがなく一般的なカメラで撮影されたかのような画像となっている．HyperOmni Vision はその構成上，一つの撮像素子で全方位の画像を得るため，画角あたりの解像度は低くなってしまふ．しかし，ビデオカメラを用いることにより容易に全方位映像が得られるという利点がある．

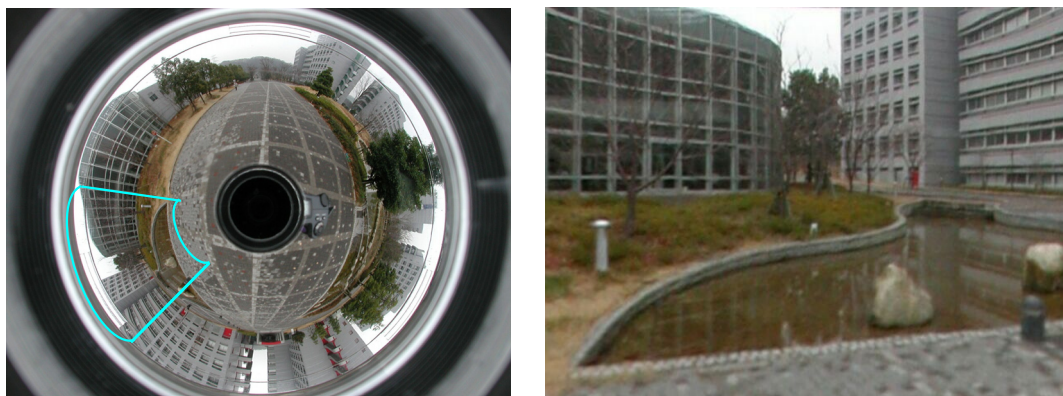


図 21 全方位画像（右）と平面透視投影画像（左）

### 2.2.2 ハードウェア機能を用いた平面透視投影画像生成

HyperOmni Vision により撮影された全方位画像をハードウェア機能を用いて高速に平面透視投影画像に変換する手法について述べる．近年の計算機の性能向上は著しく，数年前では実時間で実行不可能だった処理が，現在では実時間で実行可能といったことがよくある．しかし，テレプレゼンスシステムにおいては撮像系の進歩による画像の高解像度化も進むため，画像処理に必要とされる計算コストも増加し続けている．増加し続ける画像処理コストに対して，Onoe らは CPU（Central Processing Unit）のみで計算を行うのではなく，画像処理に適した GPU（Graphics Processing Unit）を用いて高速に画像変換を行う方式を提案した．GPU は CPU とは構造的に異なり，画像処理のようなデータ並列の計算を複数の計算ユニットによって同時に処理可能な超並列ベクトルプロセッサとして動作する．ただし，GPU は OpenGL や DirectX 等のグラフィックス命令を介して計算を実行させる必要がある．

Onoe らは，平面透視投影画像と全方位画像間の対応計算（式(2)）を各画素に対して行うのは計算コストが高いことから，生成する平面透視投影画像上に格子を設け，その各格子点に対応する全方位画像上の点の座標のみ CPU で計算し，各格子の中は GPU に計算させることで高速な変換処理を実現した．各格子内の画素に対応する全方位画像上の点の座標は，GPU のテクスチャマッピング機能による線形補間によって計算されるため，全画素の対応を CPU によって計算する

方式に比べ近似的であると言える．以下にその画像変換処理の内容を示す．

全方位画像座標  $(x, y)$  , 平面透視投影画像座標  $(u, v)$  , 3次元座標  $(X, Y, Z)$  を図 22 のように定め , 平面透視投影画像上の格子数を  $N_x \times N_y$  , 変換画像のサイズを  $W \times H$  , 視点と変換画像の間の焦点距離を  $f_h$  とする .

Step1. 任意の方向を表す回転行列  $R$  を得る .

任意の方向を ( 方位角  $\alpha$  , 仰角  $\beta$  , ロール角  $\gamma$  ) で表した場合 , 回転行列  $R$  は以下の式で表される .

$$R = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\gamma & 0 & \sin\gamma \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\gamma \end{pmatrix}$$

Step2. 変換画像上の格子点と全方位画像上の座標の対応を計算する .

変換画像上の各格子点  $(i, j)$  に対して , 以下の I ~ V の処理を実行する . ただし ,  $0 < i < N_x$  ,  $0 < j < N_y$  .

I. 格子点の変換画像上の座標を計算する .

$$u = \frac{W \times i}{N_x} - \frac{W}{2}$$

$$v = \frac{H \times j}{N_y} - \frac{H}{2}$$

II. 変換画像上の格子点の座標を保存する .

$$Pers_{i,j} = (u, v)$$

III. 変換画像上の格子点の方向を表す 3次元座標  $(X, Y, Z)$  を計算する .

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = R \begin{pmatrix} u \\ v \\ -f_h \end{pmatrix}$$

IV. 式 (2) に  $(X, Y, Z)$  を代入して全方位画像上の格子点の座標  $(x, y)$  を求める .

V. 全方位画像上の格子点の座標を保存する .  $Omn_{i,j} = (x, y)$

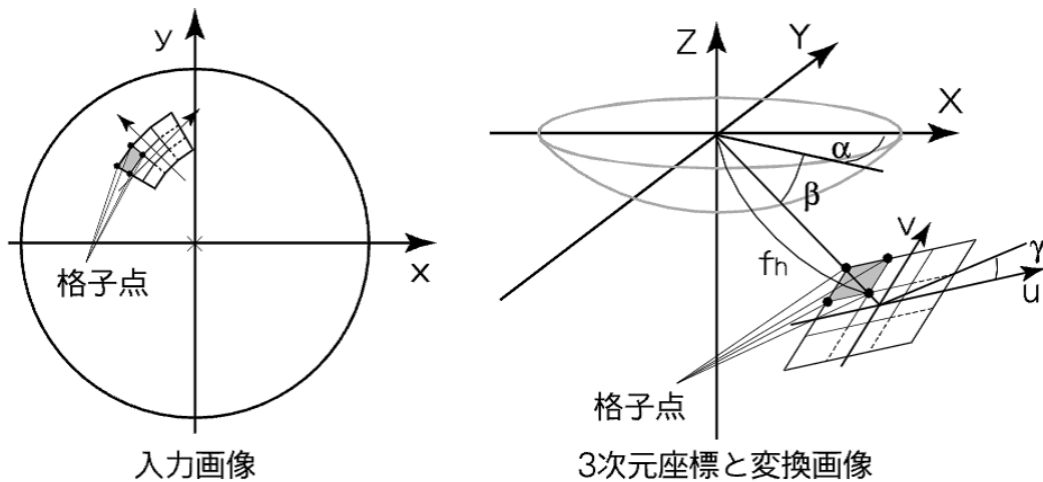


図 22 入力全方位画像（左）と 3 次元空間中の変換座標（右）

Step3. GPU のテクスチャマッピング機能を用いた Warping を実行する .

各格子点  $(i, j)$  について, 全方位画像上の  $Omni_{i,j}, Omni_{i+1,j}, Omni_{i+1,j+1}, Omni_{i,j+1}$  で囲まれる領域を変換画像上の  $Pers_{i,j}, Pers_{i+1,j}, Pers_{i+1,j+1}, Pers_{i,j+1}$  で囲まれる領域に GPU のテクスチャマッピング機能を用いて Warping する .

以上の処理により, GPU の機能を用いて高速に全方位画像から平面透視投影画像への変換を行うことができる . Onoe らは, 格子点数  $N_x \times N_y$  に対する平面透視投影画像上での幾何学的歪みの関係を示し, 生成する平面透視投影画像の大きさが  $720 \times 486$  の場合には, 格子数は  $16 \times 12$  程度で十分であるとしている . また, システムの利用者の視線方向を, HMD に取り付けられた姿勢センサにより計測し, 上記回転行列  $R$  を計算することで, その視線方向に追従した平面透視投影映像の提示を行っている .

本節で述べた画像変換方式は, 全方位画像と平面透視投影画像間の対応が与えられれば, 全方位カメラの種類に依らず適用可能である . 本研究では, 全方位映像から平面透視投影画像への変換の際には, 本節で述べた変換方式を利用する .

## 2.3 マルチキャストプロトコルによる全方位映像の配信

本節では、ネットワーク上で一对多の同報通信を可能にするマルチキャストプロトコルによる全方位映像の配信について述べる。マルチキャストプロトコルは、TCP/IP ネットワーク上で複数の相手に対して一斉に同じデータ送信するための IP の追加仕様である IP マルチキャストを行うためのプロトコルである。近年、ブロードバンド回線の普及や映像などの大容量コンテンツの人気によりネットワークの帯域は増加の一途をたどっており、効率的なデータ配信技術として IP マルチキャストは注目を集めている。

ネットワークテレプレゼンスシステムでは、通常図 23 に示すようなサーバ・クライアント型のシステム構成をとる。ある映像コンテンツを配信するサーバと、そのコンテンツを観賞するクライアント是一对多となることが想定される。従来のテレプレゼンスシステムで用いられているユニキャストプロトコルにより、データレート  $D$ [Mbps] の映像コンテンツを  $N$  台のクライアントに伝送する場合、サーバは各クライアントそれぞれに対して  $D$ [Mbps] の映像を送信する必要がある。よって、サーバに必要とされるネットワーク帯域は  $DN$ [Mbps] となる。対して、マルチキャストプロトコルによって映像を伝送した場合、サーバからパケット化され配信された映像はルータやスイッチングハブによって自動的に複製され、各クライアントに到達する。このため、サーバは 1 台分の映像を配信すればよい。そのため、必要とされるネットワーク帯域は  $D$ [Mbps] となり、ユニキャストプロトコルより効率的な映像の伝送が実現される。ただし、各クライアントに配信される映像は同一でなければ、マルチキャストプロトコルは有効ではない。全方位映像を用いたテレプレゼンスシステムでは、同一の全方位映像からあらゆる視線方向の視界画像を生成できるため、各クライアントにマルチキャストプロトコルを用いた効率的な映像伝送を行うことができる。また、サーバの処理コストの面でもクライアント 1 台分のコストですべてのクライアントに対応可能である。

全方位映像の利用者の視線方向に依存しない特徴は、マルチキャストプロトコルによる映像伝送方式と親和性が高く、利用者数の増加に対応可能なスケーラビリティが実現可能である。

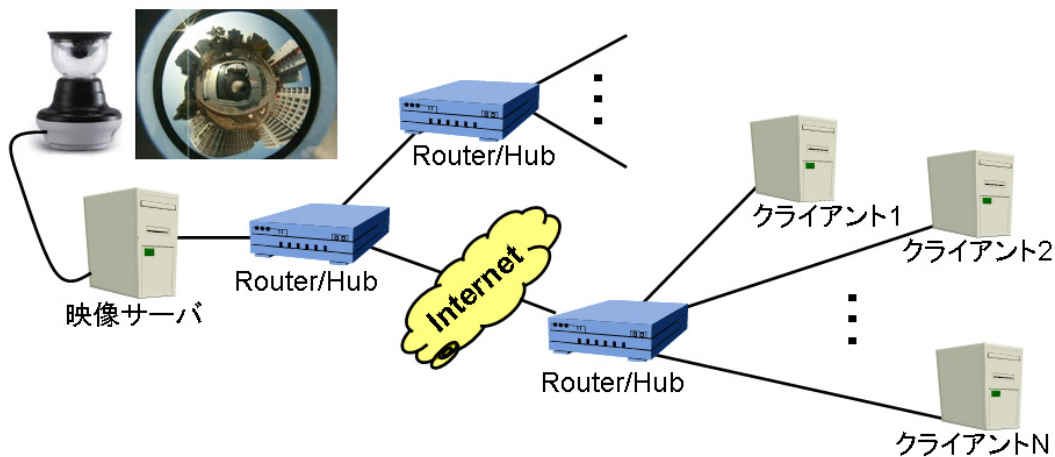


図 23 全方位映像のマルチキャスト配信

## 2.4 マルチキャストプロトコルを用いた実時間ネットワークテレプレゼンスシステム

### 2.4.1 システムの設計方針と概要

本節では、全方位映像のマルチキャスト配信による実時間ネットワークテレプレゼンスシステムの設計方針とその概要について述べる。本章で提案する実時間ネットワークテレプレゼンスシステムは、従来の全方位映像を用いたテレプレゼンスシステムの問題であった利用者数の増加に対応可能なスケーラビリティの実現及び、誰でも簡単に利用可能なユーザビリティの改善を目的とする。またこれらの改善により、従来のシステムで実現されていた機能が損なわれないよう提案システムは以下の機能にも対応する。

- ライブ型映像と蓄積型映像への対応  
実時間で取得される全方位映像だけでなく、予め蓄積しておいた全方位映像にも対応する。
- 視線追従型映像提示とビューアによる映像提示への対応  
HMD ( Head-Mounted Display ) や姿勢センサがある場合には高臨場感なインタラクティブ観賞を可能とし、それらの機器がない場合には、容易に利



用可能なビューアによるインタラクティブ観賞を実現する。

- 複数の種類の全方位カメラに対応  
用途や映像コンテンツに適した全方位カメラを選択可能とする。

本章で提案するネットワークテレプレゼンスシステムの概要を図 24 に示す。本システムでは、全方位カメラにより撮影された全方位映像を扱う。システムの利用者は、全方位カメラからの映像を実時間で配信するサーバからの全方位ビデオストリームを、Web ブラウザより起動される全方位映像ビューアにより取得する。あるいは、ネットワーク越しのサーバや PC 内に格納されている全方位映像を、同様のビューアによって取得する。この際、全方位カメラから得られた全方位映像をマルチキャストプロトコルを用いて伝送することにより、複数の利用者がシステムを利用する場合でもネットワークの負荷は増加することはない。Web ブラウザにより起動される全方位映像ビューアは、2.2.2 節で述べた GPU の機能を利用した高速な平面透視投影変換処理により視界画像を生成する。これにより、利用者は Web ブラウザ上でマウスまたはキーボード操作により全方位映像コンテンツを簡単かつインタラクティブに観賞することができる。さらに、HMD とジャイロセンサを組み合わせた視線追従型の臨場感の高い画像提示も可能である。平面透視投影変換には、複数の全方位カメラの種類に対応できるように、全方位映像と平面透視投影画像間の対応計算を複数用意した。

以下、2.4.2 節では、提案システムで扱う全方位映像の種類とその用途について述べる。2.4.3 節では、全方位映像から平面透視投影画像を生成し、利用者に提示するために用いる全方位映像ビューアについて述べる。そして、2.5 節で提案システムによる機能実証実験について述べ、2.6 節で本章で提案した全方位映像のマルチキャストによる実時間ネットワークテレプレゼンスシステムについてまとめる。

#### 2.4.2 全方位映像コンテンツ

本節では、提案システムで扱うライブ型映像と蓄積型映像について述べる。全方位映像は、ライブ中継のように実時間で映像を圧縮符号化・配信するライブ型

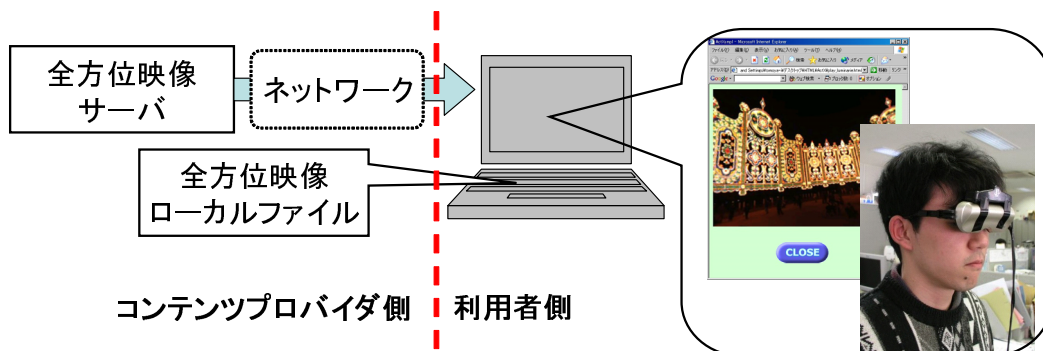


図 24 ネットワークテレプレゼンスシステムの概要

映像と、あらかじめ圧縮符号化しネットワーク越しのサーバや PC 内に保存しておく蓄積型映像に分けられる。提案システムではその両方の全方位映像を扱う。

ライブ型映像は、TV 放送のように同時に複数人が同じコンテンツを楽しむ場合に用いる。全方位カメラにより撮影された画像を瞬時に符号化・伝送し、鑑賞することができる。また、マルチキャストプロトコルを用いることによりネットワークに負荷をかけることなく多地点の複数人が同時に観賞することが可能である。

蓄積型映像は、見たいシーンを見たい時刻に見るといったオンデマンドサービスを提供する場合に用いる。また、実時間で映像を圧縮符号化し、伝送する必要がないため、圧縮符号化に時間を要する高解像度の全方位映像を扱うことが可能である。蓄積型映像では利用者それぞれが別々の映像を観賞するため、マルチキャストプロトコルによる効率的な映像伝送は行えないことに注意する必要がある。

全方位映像を撮影可能な全方位カメラには、魚眼レンズ [COH85] や曲面ミラー [YNY91, 山澤 96] によって一般的なカメラで全方位の撮影を可能にするものや、複数のカメラを外向きに配置し同時に撮影することで広視野を撮影する全方位型マルチカメラシステム [KYI<sup>+</sup>98, 棚橋 02, 池田 03] 等がある。これらの全方位カメラは、撮影対象や用途によって適切に選択する必要がある。本研究では、これら全方位カメラシステムの中でも、ミラーを用いた全方位カメラ HyperOmni Vision [山澤 96] と、複数のカメラを外向きに配置した全方位型マルチカメラシステム Ladybug [池田 03] を使用する。これらのカメラをテレプレゼンスの用途に応じて使い分ける。

### 全方位カメラ HyperOmni Vision

2.2.1 節で述べたように，HyperOmni Vision は鉛直下向きの双曲面ミラーと鉛直上向きのカメラから構成され，周囲 360 度の画像を一度に取得できる全方位カメラである．カメラ部に一般的なビデオカメラを使用することで容易に全方位映像が取得可能であり，VGA (640 × 480) 程度の解像度の映像であれば実時間で圧縮符号化を行い，遠隔に伝送することが可能である．そのため，ライブ型映像を撮影するのに適したカメラと言える．

### 全方位型マルチカメラシステム Ladybug

Ladybug は複数のカメラを外向きに配置した全方位型マルチカメラシステムである．図 25 にその外観を示す．Ladybug は画像を撮影する部分であるカメラヘッド (図 25 左) とカメラヘッドにより撮影された画像を蓄積する HDD アレイ (図 25 右) から成る．カメラヘッドは水平方向に 5 個，上方向に 1 個の計 6 個の CCD カメラを外向きに配置した構成になっており，一つの CCD カメラは横 768 画素，縦 1024 画素の高解像度な画像を撮影可能である．カメラヘッドの 6 個のカメラを同期して撮影させることにより，下方向を除く全天球の約 75 % の画像を 15fps で取得できる．カメラヘッドで撮影された画像は HDD アレイに送られ，約 20 分間の映像を蓄積できる．本研究では，池田らの手法 [池田 03] により生成された全方位映像を用いる．図 26 に池田らの手法により生成された全方位画像の例を示す．Ladybug の構成上，画像を撮影した後にオフライン処理で全方位映像を生成するため，ライブ型映像への適用はできないが，上方向を含む広範囲の高解像度映像が取得可能である．



図 25 全方位マルチカメラシステム Ladybug



図 26 池田らの手法 [池田 03] により生成された全方位画像

### 2.4.3 Web ブラウザを用いた全方位映像ビューア

本節では、提案システムにおいて利用者に対して提示する視界画像の生成や、全方位映像の受信・読み込み等の処理を行う全方位映像ビューアについて述べる。ネットワーク環境下で誰もが利用するアプリケーションに Web ブラウザがある。特に、Windows に標準で搭載されている Web ブラウザである Internet Explorer は JAVA アプレットや ActiveX によるプログラムを実行可能で、利用者に対し Web サイト上でインタラクティブなコンテンツを提供することを可能にしている。また、JAVA アプレットや ActiveX プログラムは自動インストール機能により利用者に負荷をかけることなく提供可能という特徴がある。

図 27 に本システムで用いる全方位映像ビューアの構造を示す。全方位映像ビューアはグラフィックスカード上の GPU の機能を使用するため、ActiveX プログラムとして実装を行う。全方位映像ビューアは Web ブラウザにより起動され、全方位映像を平面透視投影画像に変換した後、Web サイト上にその画像を提示する。以下にその処理手順を述べる。

Step1. 利用者が全方位映像コンテンツを提供している Web サイトにアクセスすると、Web ブラウザにより全方位映像ビューアが起動される。全方位映像ビューアは、受信する全方位映像のファイルパスやその全方位映像を撮影した全方位カメラの内部パラメータ等の情報を、自身が起動される Web サイトから読み取る。このビューアに必要なパラメータはコンテンツプロバイダが作成しておけばよいので、利用者はパラメータを意識することなくコンテンツを観賞可能となる。

Step2. ビューアは DirectShow の機能を使用し、ビューアに入力された全方位映像のファイルパスの情報を基にサーバにアクセスし、全方位映像の受信を開始する。全方位映像が利用者の PC 内に格納されているのであれば、その PC 内から読み出しを始める。映像の入力に DirectShow を使用しているため、AVI や MPEG、WMV 等様々なファイルフォーマットに対応できる。そして、入力される全方位映像から 1 フレームずつ全方位画像を取得する。

Step3. 順次取得される全方位画像を OpenGL の機能を使用してグラフィックス

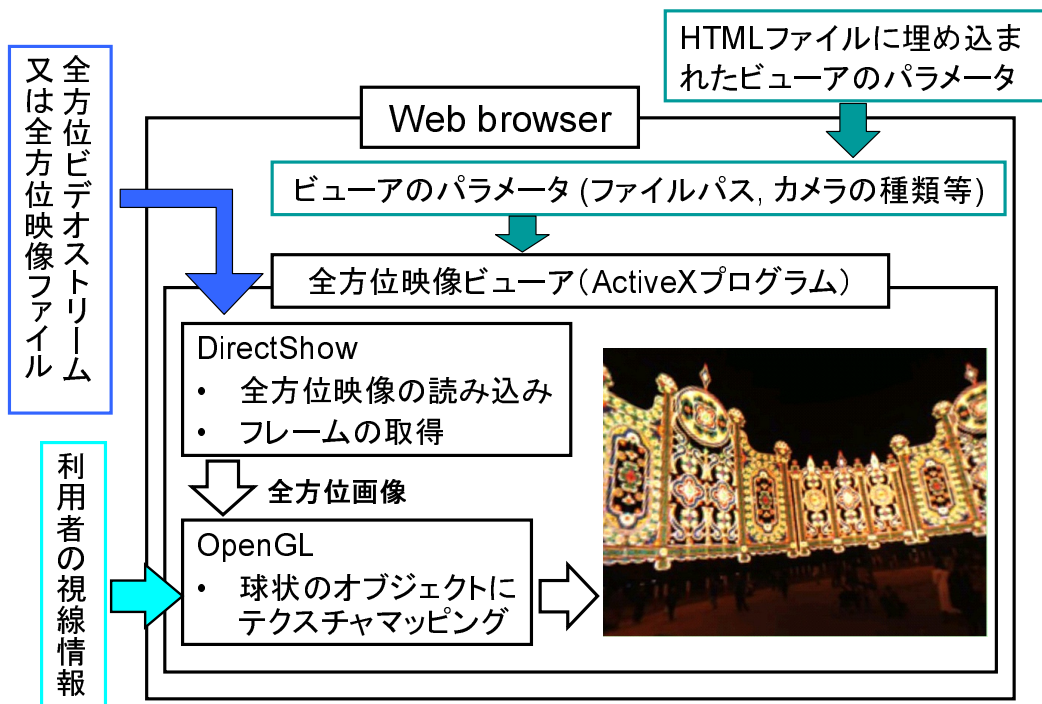


図 27 全方位映像ビューアの構造

カードのテクスチャメモリへ転送する．そして，2.2.2 節で述べた高速な平面透視投影変換処理により，利用者に提示する視界画像を生成する．画像生成の際の視線方向はマウスやキーボード，ジャイロセンサにより得られる情報により設定される．

Step4. 最後に，レンダリングされた視界画像を Web サイト上に表示する．また，利用者が HMD を使用して全方位映像を観賞するときには，HTML ファイルにフルスクリーンで画像提示を行うパラメータを記述しておくことで，全方位映像がフルスクリーンで提示される．

## 2.5 ネットワークテレプレゼンス実験

以上に述べたネットワークテレプレゼンスシステムを実装し，蓄積型映像とライブ型映像を用いた機能実証実験を行った．蓄積型映像を用いた実験では，提案

システムのオンデマンドサービスへの適用を想定し、全方位型マルチカメラシステムにより撮影された高解像度の全方位映像をネットワークを介してインタラクティブに観賞することを目的とした。一方、ライブ型映像を用いた実験では、提案システムの TV 中継への適用を想定し、全方位カメラ等を搭載した車両の走行中の全方位ライブ映像を車両から直接マルチキャスト配信することによって複数人が同時に映像を観賞することを目的とした。また本システムで用いる全方位映像ビューアにより、Web ブラウザ上でマウス操作により全方位映像をインタラクティブに観賞できることと、HMD と姿勢センサを組み合わせた視線追従型映像提示による高臨場感の映像観賞が可能であることを確認する。

#### 2.5.1 高解像度な蓄積型全方位映像を用いたテレプレゼンス

本実験では、提案システムを全方位映像のオンデマンドサービスへ適用した場合にも利用可能であることを確認する。先に述べたように、オンデマンドサービスでは実時間で全方位映像の圧縮符号化を行う必要がないため、高解像度の全方位映像を予め圧縮符号化しておくことが可能である。本実験では、全方位型マルチカメラシステム Ladybug により得られた全方位映像を予め圧縮符号化してネットワーク (100Mbps) 内の全方位映像サーバに格納しておき、その全方位映像を利用者がオンデマンドで観賞する。利用者は観賞用 PC (表 1 参照) で、コンテンツプロバイダが用意した Web サイトにアクセスすることで、映像の観賞ができる。実験に使用する全方位映像は、観賞用 PC の性能の制約からその解像度を  $1344 \times 672$  に、フレームレートを 30fps (15fps の映像の各フレームを 2 回提示) に設定した。また映像のフォーマットは、映像部には復号化処理が高速である理由から MPEG1 を、音声部には MPEG1 Layer2 を用いた。利用者は全方位映像コンテンツを Web ブラウザや HMD により観賞する。

図 28 に観賞用 PC の Web ブラウザ画面を示す。マウスで画面上をドラッグ操作することにより自由に視線方向を変更でき、簡単に全方位映像コンテンツを観賞可能であることを確認した。高解像度の全方位映像であるが、本システムで用いる全方位映像ビューアの高速な平面透視投影変換処理により、30fps でのインタラクティブな映像観賞が可能であった。また、映像と音声は全方位映像ビューア

表 1 蓄積型映像観賞用 PC の詳細

CPU	Pentium4 2GHz
メモリ	512MB
グラフィックスカード	ATI RADEON9700pro
OS	WindowsXP

内の DirectShow 機能により同期して再生された。利用者の一人が映像観賞中，別の利用者が映像を始めから観賞し始めたが，システムは問題なく動作した。ただし，ユニキャストプロトコルによる映像伝送のため，利用者が多数になるとサーバの負荷が増加し，フレームが欠落する現象が発生した。

より臨場感の高い観賞方法として，図 29 に示すような HMD に利用者の視線方向を検出するためのジャイロセンサを取り付けた視線追従型の画像提示も行った。使用したジャイロセンサは INTERSENSE 社の InterTrax2 である。このジャイロセンサは姿勢に関する情報を最大 256Hz で取得可能である。マウス操作による観賞を行う場合と同様に，視線追従型映像提示用の Web サイトにアクセスすると，HTML に記述されているパラメータにより，HMD で表示するために映像がフルスクリーンで提示されるモードでビューアが動作する。姿勢センサ付き HMD を装着し観賞を行い，視線方向の変更からその方向の画像が提示されるまでの時間遅延は約 17ms であり，視線方向を自由に変えることができた。本実験により，提案システムがオンデマンドサービスへの適用が可能であることを示した。また，マウス操作による簡易な全方位映像観賞と HMD による視線追従型映像観賞の両方が利用可能であることを確認した。



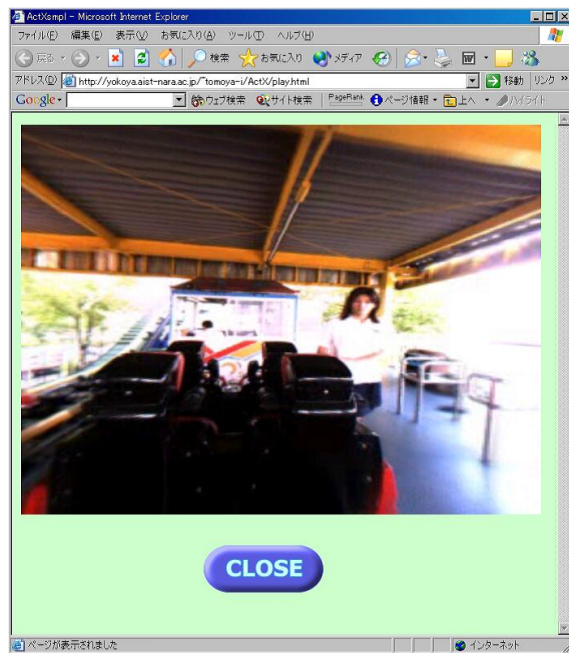
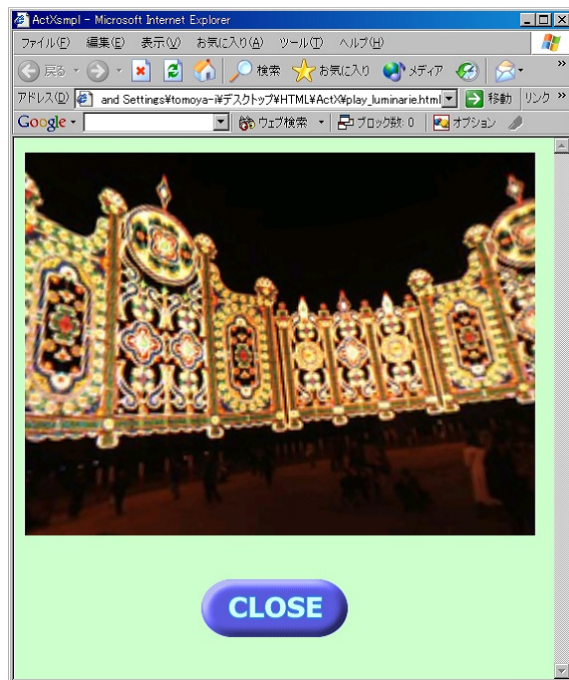


図 28 観賞用 PC の Web ブラウザ画面



図 29 ジャイロセンサを取り付けた HMD

### 2.5.2 車載全方位カメラからの全方位ライブ映像を用いたテレプレゼンス

本実験では，提案システムの TV 中継への適用を想定し，全方位カメラ等を搭載した車両の走行中の全方位ライブ映像を車両から直接マルチキャスト配信することによって複数人が同時に映像を観賞可能であることを確認する．図 30 に示すように，全方位カメラ HyperOmni Vision を搭載した車両，全方位映像を取得・圧縮符号化・マルチキャスト配信する PC，全方位映像ビューア，またそれらをつなぐ無線・有線ネットワークからなるシステムを構築した．全方位カメラを搭載した車両は，本学屋外の道路を走行しながら全方位カメラで周囲 360 度を撮影し，i.Link で車内の画像取得圧縮符号化・マルチキャスト配信用 PC に全方位映像を送る．画像取得圧縮符号化・マルチキャスト配信用 PC では，全方位映像（解像度:640 × 480，フレームレート:24fps）を WindowsMediaEncoder[Cor02] によって WindowsMedia 形式（ビットレート:832Kbps）に実時間で圧縮符号化する．符号化された全方位映像は，マルチキャストプロトコルによって無線ネットワークに配信される．表 2 及び図 31,32 に画像取得車両システムの構成と外観を示す．ネットワークに配信された全方位映像は観賞用 PC 上で動作する全方位映像ビューアを用いて観賞することができる．ユニキャストではなくマルチキャストで配信することにより受信するシステムが多数あったとしてもネットワークの負荷は増加しない．ネットワークがマルチキャストプロトコルに対応していれば，多数の利

表 2 画像取得車両のシステム構成

全方位カメラ	SONY DCR-TRV900 + アコウル 双曲面ミラー
画像取得圧縮符号化・ マルチキャスト配信用 PC	Pentium4 2.53GHz Memory 1GB WindowsXP
無線ネットワーク	IEEE802.11g
車両	日産 ELGRAND

用者がビューアにより全方位映像を受信することが可能である。

本実験では、5台の観賞用 PC で同時に全方位映像の観賞を行い、全ての観賞用 PC でそれぞれ任意の方向を観賞することが可能であることを確認した。それらの PC の全方位映像ビューアの画面例を図 33 に示す。Web ブラウザ上の映像は送信した全方位映像と同様の 24fps で提示され、全方位カメラによる環境の撮影とその映像が提示されるまでには約 10 秒の時間遅延があった。この時間遅延の主な原因は、全方位カメラからの映像を圧縮符号化するためのバッファリングによるものと、全方位映像ビューアが圧縮符号化された全方位映像を復号化するためのバッファリングによるものである。また、ネットワーク負荷は 832Kbps で一定であり、映像を受信する観賞用 PC の数が増えてもネットワークの負荷が増加しないことを確認した。無線ネットワークは有線のそれに比べ、利用可能な帯域が狭く不安定でありユニキャストによる映像配信では複数の利用者が同時に映像を観賞することは困難であるが、映像のマルチキャスト配信により観賞用 PC が複数であった場合にも安定して観賞を行うことができた。

## 全方位映像ビューア

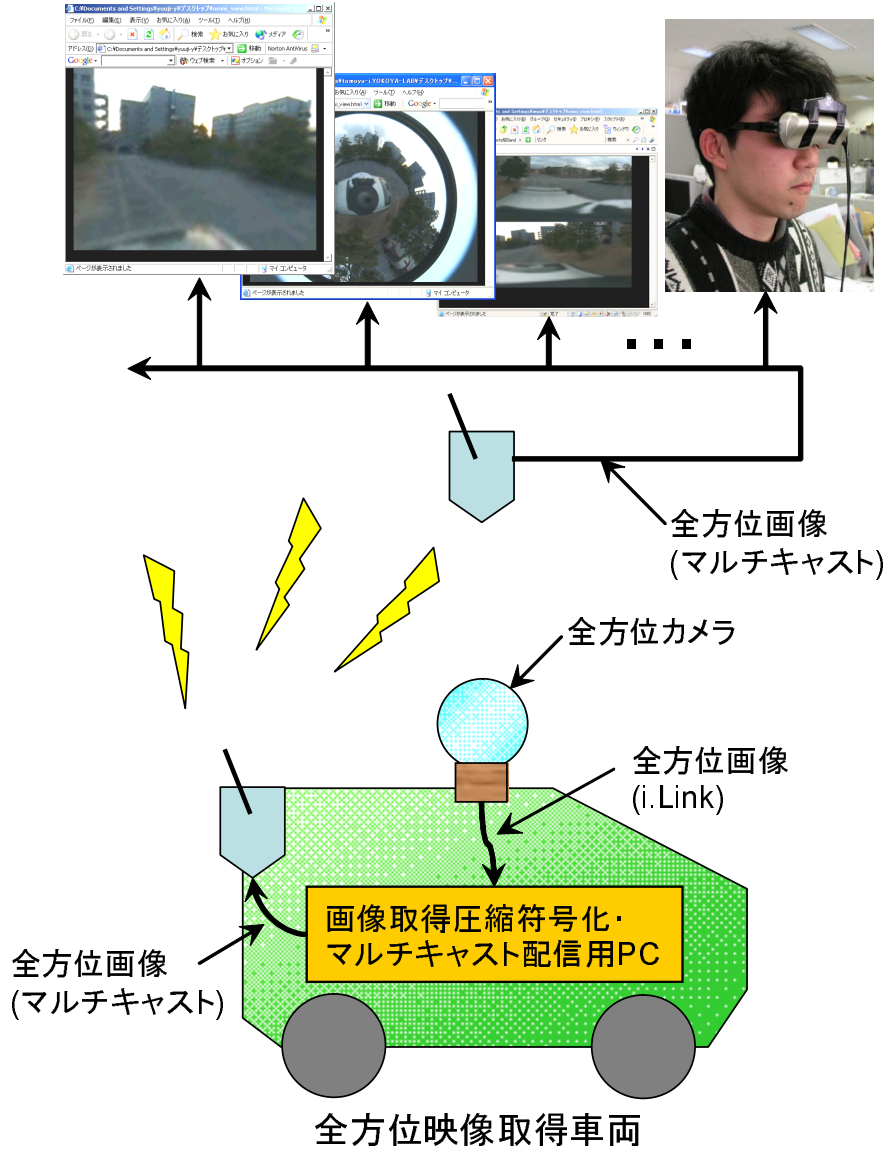


図 30 ライブ型映像のマルチキャスト配信実験のシステム構成



図 31 全方位映像取得車両の外観



図 32 全方位映像取得車両の内部

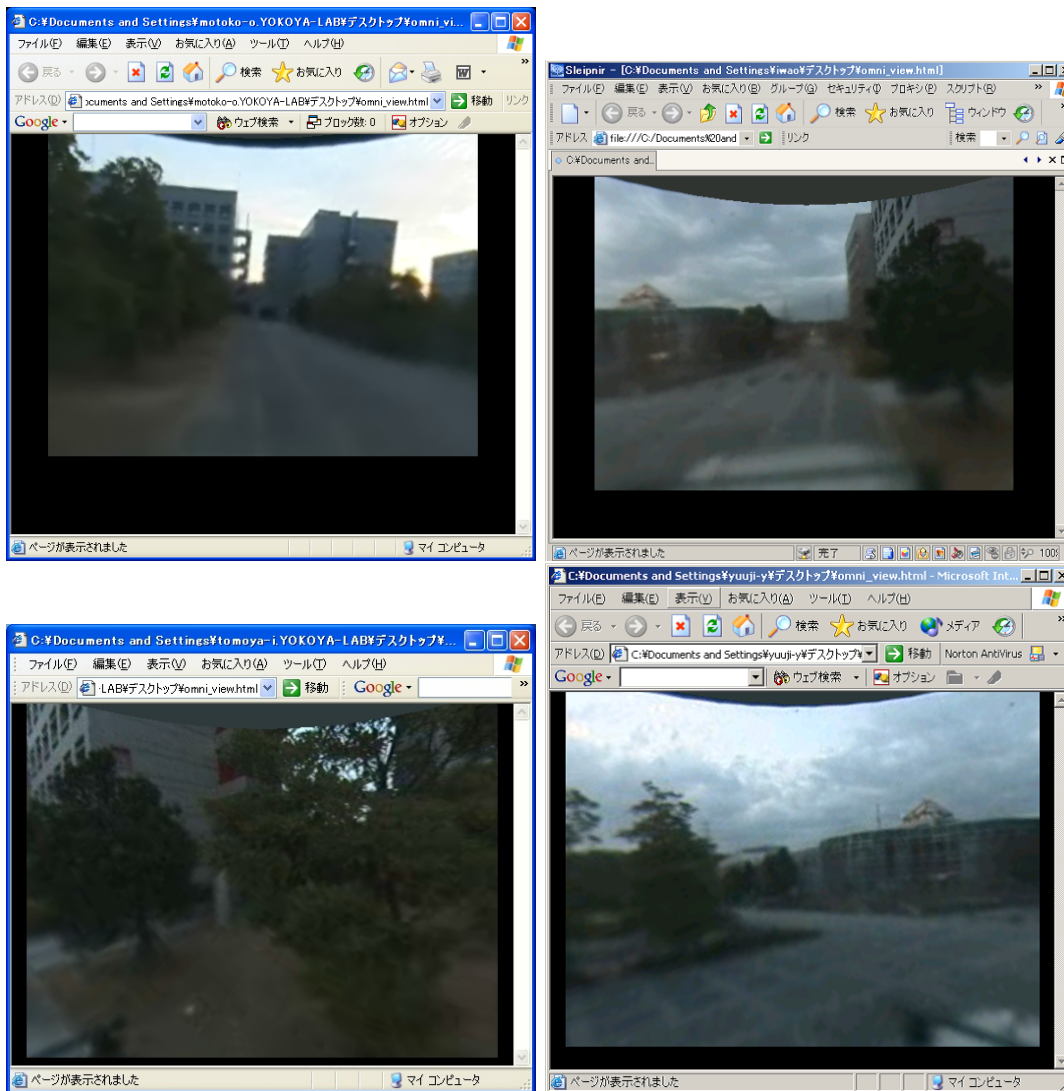


図 33 利用者それぞれの全方位映像ビューアの画面例

## 2.6 結言

本章では、全方位映像のマルチキャストによって、複数の利用者がシステムを同時に使用する場合にも対応可能なハイスケーラビリティを有する実時間ネットワークテレプレゼンスシステムを提案した。また、ユーザビリティを考慮した全方位映像ビューアにより、利用者が Web ブラウザ上で簡単に全方位映像コンテンツを観賞可能とした。さらに、TV 中継のように実時間配信されたライブ映像のみならず、オンデマンドで映像を提示するテレプレゼンスにも対応可能である。

蓄積型映像を用いた実験では、利用者はオンデマンドで高解像度な全方位映像コンテンツを観賞することが可能であることを確認した。観賞用 PC の性能による制約から全方位映像の解像度を  $1344 \times 674$  に落とす必要があったが、Ladybug から池田らの手法により得られる全方位映像の最大解像度は  $3840 \times 1920$  であり、今後計算機の性能向上によりこのような高解像度の全方位映像も扱えるようになると考える。

ライブ型映像を用いた実験では、多数の利用者が実時間でマルチキャスト配信される全方位ライブ映像を同時に観賞し、利用者数が増加した場合にもネットワークの負荷は一定であることを確認した。全方位映像の取得に用いた全方位カメラが NTSC 程度の解像度であったため、平面透視投影変換した画像に粗さが目立っていた。今後計算機の高性能化により、Ladybug の映像そのものや、HD カメラを用いた HyperOmni Vision の映像を実時間で圧縮符号化可能になり、ライブ型映像においても高解像度の映像観賞が可能になると考える。また、ライブ型映像の実験では、全方位カメラによる撮影からその映像の提示までに約 10 秒の時間遅延があった。しかし、この時間遅延は利用者のインタラクティブな映像観賞には影響を与えない。時間遅延の主な原因は映像圧縮符号・復号化の際のバッファリングによるものであり、本システムを遠隔地の相手とのコミュニケーション等の用途に適用する場合には、より時間遅延の少ない映像圧縮符号化方式の利用も検討する必要がある。

### 3. 多視点全方位映像を用いたハイスケーラブル自由視点テレプレゼンス

#### 3.1 序言

本章では，前章で提案したハイスケーラブルな実時間ネットワークテレプレゼンスシステムを基に，多視点全方位映像からの実時間自由視点画像生成処理を加え，1.2節で述べた要求機能 A～D を全て満たす実時間ネットワークテレプレゼンスシステムを提案する．

前章での全方位映像のマルチキャスト配信による実時間ネットワークテレプレゼンスシステムの実現により，要求機能 A,B,D 及び C の視線追従機能を満たすことができた．このテレプレゼンスシステムに，自由視点画像生成機能を追加することで要求機能の全てを満たすシステムが実現できる．前章のシステムに自由視点画像生成機能を追加するには，全方位の動的環境に対応可能かつ，実時間での画像生成が可能な手法を用いる必要がある．しかし，1.3節で述べた通り，従来の自由視点画像生成手法をそのまま適用することはできない．そこで本研究では，多地点に配置した全方位カメラからの多視点全方位映像を用いて実時間で自由視点画像を生成する手法を提案する．提案手法では，撮影環境を静的領域と動的領域に分離し，それぞれの領域に適した自由視点画像生成手法を用いる．その際，計算コストの高い動的領域の自由視点画像生成に対しては，動物体の存在領域推定を併用した高速な画像生成処理を行う．そして，静的領域と動的領域の自由視点画像を合成することで最終的な自由視点全方位画像を得る．

以下では，想定する提案システムの利用環境と，環境に配置された全方位カメラの内部・外部パラメータのキャリブレーション法について述べ，本章で提案する自由視点画像生成手法について詳述し，シミュレーションによる画質評価により全方位カメラの配置について検討する．そして，その自由視点画像生成処理を用いたハイスケーラブル自由視点テレプレゼンスシステムを提案する．実験により，要求機能 A～D の機能を全て満たすことを確認し，被験者による主観評価により提案システムの有効性を確認する．



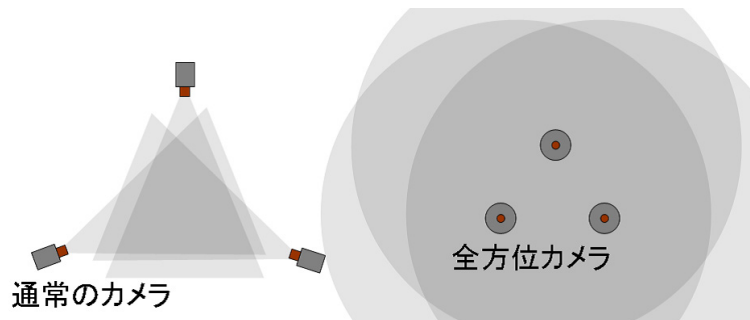


図 34 カメラ配置別の仮想化可能な空間範囲

### 3.2 想定するシステムの利用環境

本節では、本研究で想定するテレプレゼンスシステムの利用環境について述べる。本研究では、遠隔の観光地や名所の風景をインターネットを介してインタラクティブに観賞するようなアプリケーションを想定し、その用途に適したシステムの構築を目指している。このような用途に対しては、Kanadeら [KNR97] や Saitoら [SBK03] が採用している通常のカメラ群が特定領域を集中して撮影するようなカメラ配置（図 34 左参照）ではなく、複数の全方位カメラによって遠隔の環境全体を撮影するような配置が望ましい（図 34 右参照）。複数の全方位カメラにより環境全体を撮影することで、全方位に対応した自由視点画像生成が可能になる。

提案システムでは、環境撮影用のカメラ群をその観光地や名所での主要な観賞スポットに配置し、利用者はそのスポット単位での観賞を行うことを想定する（図 35 参照）。利用者は、遠隔の室内環境中で HMD と位置・姿勢センサを装着し、視点・視線に追従した映像観賞を行う。そのため、スポット内での利用者の移動可能範囲は約 2m 四方程度とする。この利用者の自由視点位置を包含する多角形位置にある複数の全方位カメラからの画像を利用して自由視点画像を生成し提示する。以下、本研究では二次元的な視点移動を可能にするために最低必要な 3 台の全方位カメラを利用することを想定し、提案する自由視点画像生成手法とシステムについて述べる。また、本研究は各観賞スポットにおけるインタラクティブな映像提示技術に焦点を当て、各観賞スポット間の移動は、従来の蓄積再生型の映像提示手法等の利用により行われることを想定する。

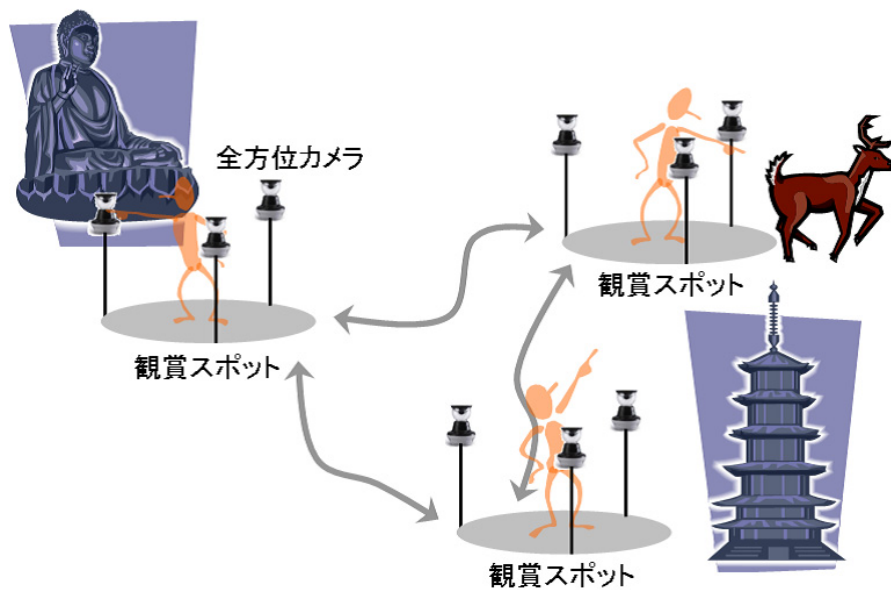


図 35 システムの利用環境

### 3.3 全方位カメラの内部・外部パラメータのキャリブレーション

本節では、環境中に配置した全方位カメラの内部・外部パラメータの推定方法について述べる。提案する自由視点画像生成手法は、撮影環境（世界座標系）に対する各全方位カメラの位置・姿勢情報（外部パラメータ）と、撮影環境中の点の全方位画像上への投影変換に必要な内部パラメータを要する。本研究では図 36 に示すように、環境中の 3 次元位置が既知の基準点  $P_i = [X_i, Y_i, Z_i](i = 1, \dots, I)$ （以下、マーカと呼ぶ）を用いて全方位画像上の主点座標（Principal point） $C = [u_c, v_c]$  と焦点距離  $f$ ，世界座標系に対するカメラ座標系の位置  $T_M = [t_x, t_y, t_z]$ ・姿勢  $R_M = [r_x, r_y, r_z]$  の 9 パラメータ（表 3 参照）を推定する。ここで推定した内部・外部パラメータは後述の自由視点画像の生成結果に大きな影響を与えるため、精度良く推定する必要がある。

本研究では、根岸ら [根岸 04] と同様のアプローチにより、全方位カメラの各パラメータ ( $C, f, T_M, R_M$ ) を推定する。その際、ミラーの内側の焦点はカメラ部の光軸上であり、カメラ部の投影中心はミラーの外側の焦点と一致すると仮定する。また、ミラーの形状パラメータ  $a, b, c$  は既知かつ誤差が含まれないものとす

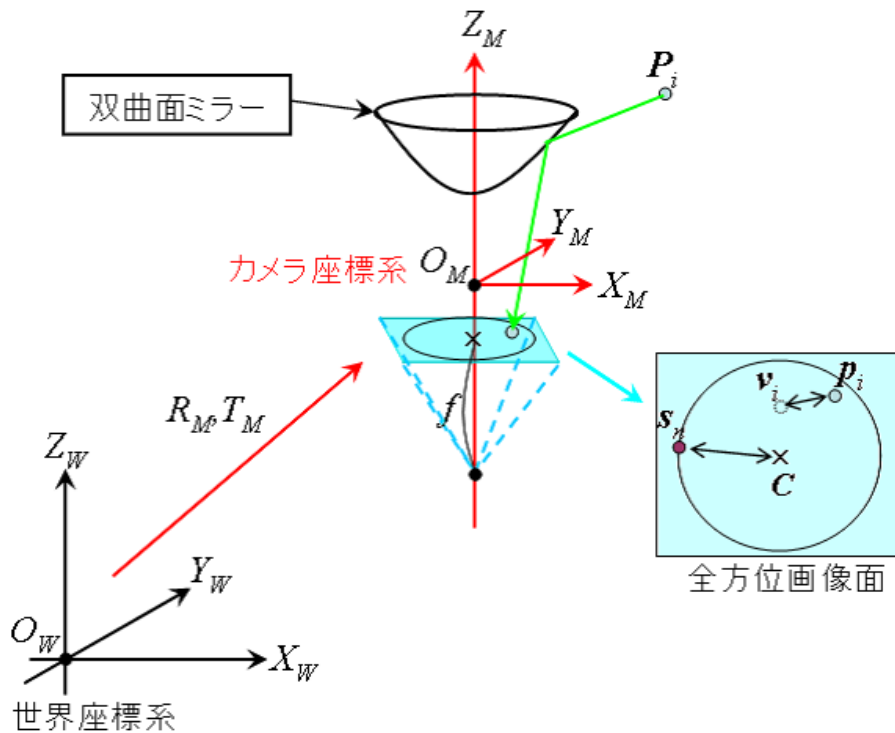


図 36 世界座標系と推定する全方位カメラの内部・外部パラメータの関係

表 3 推定する全方位カメラの内部・外部パラメータ

全方位画像上の主点座標	$C = [u_c, v_c]$
カメラの焦点距離	$f$
カメラ座標系の位置	$T_M = [t_x, t_y, t_z]$
カメラ座標系の姿勢	$R_M = [r_x, r_y, r_z]$

る．この仮定の下で，3次元位置が既知のマーカの再投影誤差を最小化するアプローチにより，全方位カメラの各パラメータを推定する．

まず，主点座標  $C$  を画像上に投影されたミラーの縁を利用することで推定する．根岸らはキャプチャボードの影響による画像のアスペクト比の変化に対応するために，ミラーの縁と楕円方程式をフィッティングすることで  $C$  を推定している．本研究では HyperOmni Vision のカメラ部にアスペクト比が 1.0 のデジタル

カメラを使用することを想定し， $C$  の推定には円方程式を用いる．この場合，ミラーの縁上の点を  $s_n = [u_n, v_n]$  ( $n = 1, \dots, N$ )， $h$  を円の半径とすると，これらの点と円方程式の誤差関数  $E_C$  は以下の式で定義できる．そして， $E_C$  の非線形最小化処理によって  $C$  を推定する．

$$E_C = \sum_{n=1}^N (\|s_n - C\| - h)^2 \quad (3)$$

次に，上で推定した主点座標  $C$  と暫定的に設定した 7 パラメータ ( $f, T_M, R_M$ ) で 3 次元位置が既知のマーカ  $P_i$  が式 (2) により投影される画像上の点  $p_i$  と，実際に画像上で検出された点  $v_i$  との二乗誤差  $E_p$  (式 4) を非線形最小化することで，全方位カメラの焦点距離と位置・姿勢を推定する．このとき，1 枚のキャリブレーションボードを用いて推定を行った場合，全方位画像中の特定方向にしかマーカが写らないため，推定結果に偏りを生じる．さらに，同一平面にしかマーカを配置できないため，精度の良い推定が困難となる．そこで，環境中の特徴的な点 (建物や机の角等) を，3 次元位置計測器 (トータルステーション) により偏りなく計測することで，空間的に全方向に分布したマーカ集合を得る．これにより，推定結果の偏りを防止し，さらに精度の良い推定が可能となる．図 37 に，環境中のマーカと推定されたカメラの位置・姿勢の例を示す．図中の点がキャリブレーションに使用したマーカの位置，錘台が全方位カメラの位置・姿勢を示している．

$$E_p = \sum_{i=1}^I \|v_i - p_i\|^2 \quad (4)$$

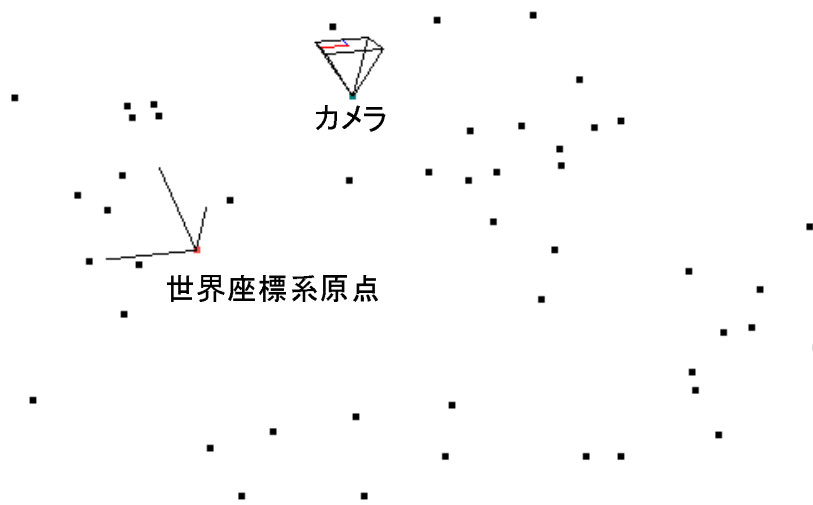


図 37 環境中のマーカと推定されたカメラの位置・姿勢

### 3.4 多視点全方位映像からの実時間自由視点画像生成

本節では，環境中に配置された複数の全方位カメラから得られる全方位画像からの自由視点画像生成手法について述べる．提案手法では，多地点に配置されたカメラのうち，自由視点位置を取り囲むような三角形をなすことができる三つのカメラを選択する．これは，後述する Morphing に基づく自由視点画像生成の際に，2次元平面的な視点移動を行うために最低限必要なカメラ台数である．そして，その3台のカメラから成る三角形内部の平面上を自由視点位置の移動可能範囲とする．また，撮影環境の大部分は静的であり，動的領域は全方位画像中の小領域であると仮定する．加えて，各全方位カメラは環境に固定され，互いに環境を重複して撮影できるよう設置されており，同期済みであるとする．その環境に設置された各全方位カメラの位置・姿勢は3.3節で述べた方法により推定する．

図38に自由視点画像生成手法の処理の流れを示す．まず，3地点に配置した全方位カメラからの同時刻の全方位画像を取得する．取得した各全方位画像に対し背景差分処理を行い，画像を静的領域と動的領域に分離する．静的領域と動的領域を分離することで，時間的に変化の少ない静的領域に適した高速な自由視点画

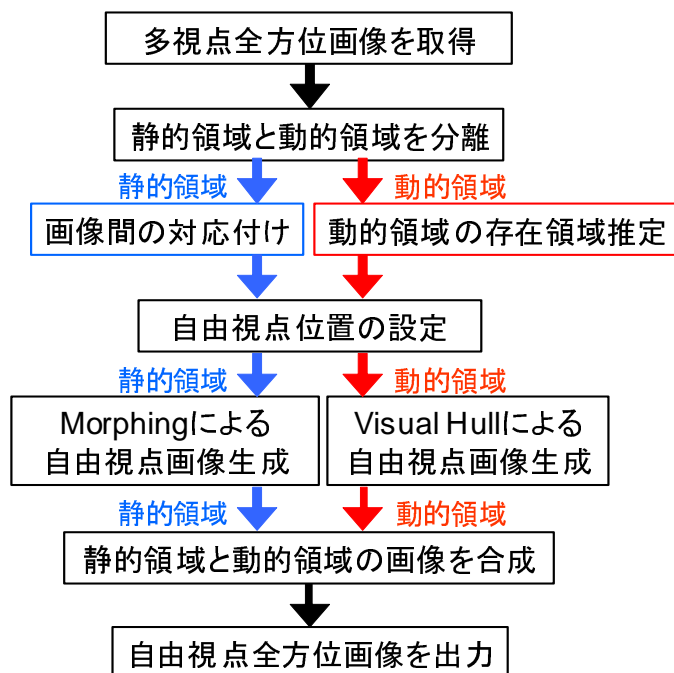


図 38 多視点全方位映像から自由視点画像生成処理の流れ

像生成手法を適用可能にする．また，背景差分処理には長時間にわたって背景差分処理を続けられるよう，緩やかな照明変化や短い周期の輝度変化を考慮した背景差分手法 [YY03] を用いる．

静的領域と動的領域の分離後，それぞれの領域に対して自由視点画像生成を行う．まず，静的領域に対しては，画像間の対応点に基づいた Morphing [冨手 02] により自由視点画像を生成する．動的領域に対する自由視点画像生成には Visual Hull を用いる．一般に Visual Hull はボクセルによって表現されることが多いが，撮影される空間が広い場合には計算量とデータ量が膨大となる．そこで本研究では，Image-based Visual Hull [MBR+00] による Visual Hull の計算を行う．さらに，動的領域の存在領域推定を併用することで高速な処理を実現する．それぞれの領域の自由視点画像を生成後，それらを合成することで最終的な自由視点画像とする．以下では，各処理について詳述する．

### 3.4.1 背景差分処理による静的領域と動的領域の分離

3 地点の全方位カメラで取得された画像それぞれに対して、静的領域と動的領域を分離する。この処理のための基本的な戦略としては、予め動的領域の存在しない背景画像を撮影しておき、背景画像と入力された画像との差分によって動的領域を検出し、分離することが考えられる。ただし、この戦略では、照明変化やノイズに弱く、長時間にわたって動的領域を検出するには困難である。そこで本研究では、蛍光灯やディスプレイのフリッカー、緩やかな照明変化、カメラに依存したノイズ等を考慮した背景画像の更新を伴う背景差分手法 [YY03] を用いる。

この手法は、時系列的に入力される画像の各画素について輝度の平均値や標準偏差値を保持し、現フレームの輝度値が与えられた背景画素の輝度値のモデルの範囲内であれば背景と見做し、それ以外の場合には動的領域と見做す。画像中の各背景画素の輝度  $I$  は式 (5) でモデル化される。

$$I = \bar{I} + \sigma \sin(2\pi\omega t) + k\xi \quad (5)$$

$\bar{I}$  は輝度の時間平均、 $\sigma$  は輝度の振幅、 $\omega$  は輝度の周波数、 $k$  は  $-1 \sim +1$  の値をとる係数、 $\xi$  はカメラのみに依存した雑音を表す。 $\sigma \sin(2\pi\omega t)$  の項は蛍光灯や CRT ディスプレイ等のフリッカー、木々の揺らぎなどの影響を表し、 $k\xi$  の項はカメラに依存するノイズを表す。このとき背景画素の輝度値  $I$  は以下の範囲をとり得る。

$$\bar{I} - \sigma - \xi \leq I \leq \bar{I} + \sigma + \xi \quad (6)$$

$\bar{I}$  と  $\sigma$  は、時系列に入力される画像の輝度の平均値と標準偏差値として計算でき、 $\xi$  は動物体が全く存在しない環境を撮影したときの標準偏差値で計算できる。入力画像の各画素の輝度値が式 (6) の範囲内であればその画素は静的領域と見做し、そうでなければ動的領域と見做す。また、入力画像の RGB 値それぞれに対して、動的領域の判定処理を行い、3 色中 1 色でも動的領域と判断されれば、その画素は動的領域と判定する。それぞれの領域は、以降の処理のために別々の画像として保持しておく。

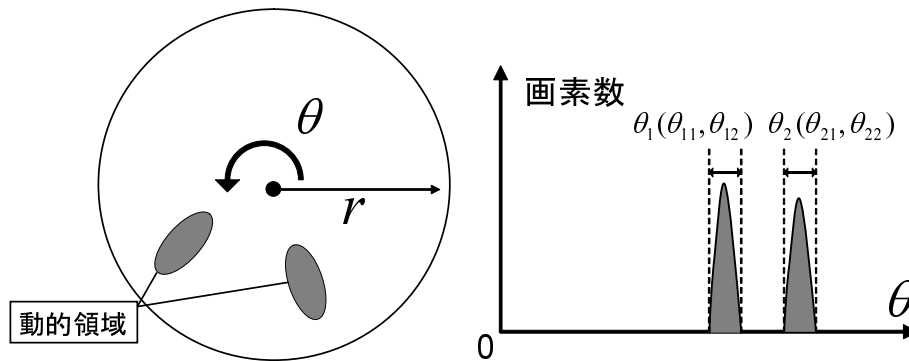


図 39 極座標変換によるヒストグラムの作成

### 3.4.2 動的領域の存在領域推定

本節では、動的領域のみが含まれる全方位画像を基に、動的領域の存在領域を推定する処理について述べる。この処理によって算出された存在領域は、後述する Image-based Visual Hull の高速化に使用される。存在領域の推定には、森田らの手法 [森田 05] を用いる。以下に、その処理を概説する。

動的領域のみが含まれる全方位画像に対して直交座標から画像の中心を原点とする極座標  $(r, \theta)$  に変換し、 $\theta$  軸方向に作成したヒストグラムから連続したある閾値より大きい値を持つ部分の画素領域の方位角範囲  $\theta_n(\theta_{n1}, \theta_{n2})$  を求める (図 39 参照)。この際、ノイズ除去処理として、一定以上の幅を持たない方位角範囲は除去する。

得られた方位角範囲から動的領域の存在領域を推定する。図 40 より、動的領域の存在領域は方位角範囲  $\theta_n(\theta_{n1}, \theta_{n2})$  の各ベクトルの交点から計算できる。各ベクトルの交点座標は、世界座標系でのカメラの位置ベクトルと方位角範囲の両端の角度に対応する単位ベクトル及び媒介変数を用いて以下の式で求めることができる。

$$(\text{交点座標}) = (\text{カメラの座標値}) + (\text{単位ベクトル}) \times (\text{媒介変数})$$

よって 2 つの直線の交点座標は、直線を延ばしたカメラの位置ベクトルをそれぞれ  $\vec{S}_1(S_{1x}, S_{1y})$ ,  $\vec{S}_2(S_{2x}, S_{2y})$ , 直線の単位ベクトルをそれぞれ  $\vec{V}_1(V_{1x}, V_{1y})$ ,  $\vec{V}_2(V_{2x}, V_{2y})$ , 媒介変数をそれぞれ  $M_1$ ,  $M_2$  とすると、以下の式で表せる。



$$\begin{aligned}
 (\text{交点の } x \text{ 座標}) &= S_{1x} + V_{1x} \times M_1 \\
 &= S_{2x} + V_{2x} \times M_2
 \end{aligned} \tag{7}$$

$$\begin{aligned}
 (\text{交点の } y \text{ 座標}) &= S_{1y} + V_{1y} \times M_1 \\
 &= S_{2y} + V_{2y} \times M_2
 \end{aligned} \tag{8}$$

式 (7),(8) より，各媒介変数は以下の式で求めることができる（図 41 参照）．

$$M_1 = \frac{(S_{2x} - S_{1x}) \times V_{2y} - (S_{2y} - S_{1y}) \times V_{2x}}{V_{1x} \times V_{2y} - V_{1y} \times V_{2x}}$$

$$M_2 = \frac{(S_{1x} - S_{2x}) \times V_{1y} - (S_{1y} - S_{2y}) \times V_{1x}}{V_{2x} \times V_{1y} - V_{2y} \times V_{1x}}$$

動的領域の存在領域は，交点座標の組として保持し，以降の処理に用いる．

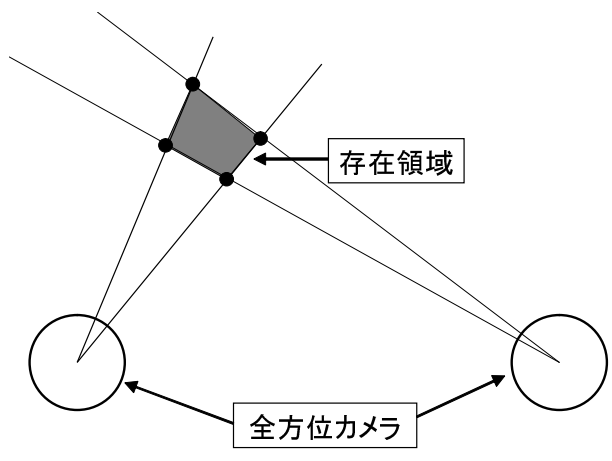


図 40 推定される存在領域

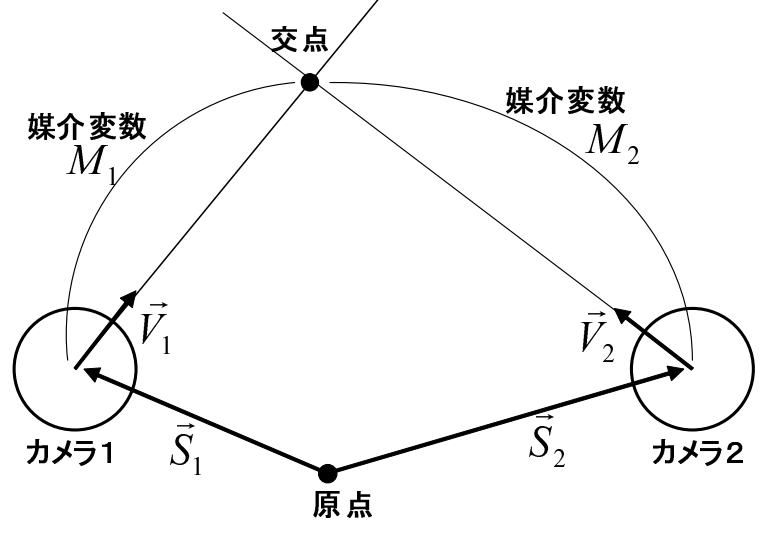


図 41 直線の媒介変数  $M$

### 3.4.3 静的領域に対する Morphing による自由視点画像生成

3.4.1 節で生成された静的領域のみが含まれる画像に対して，Morphing を用いた自由視点画像生成手法 [富手 02] を適用する．この手法は，複数の全方位画像間の対応点を必要とするが，環境中の静的領域は長時間変化することがないと仮定し，予め対応点を与えておく．

多地点全方位画像間の対応を与えることにより環境の疎な 3 次元形状を復元し，さらに GPU ( Graphics Processing Unit ) の機能を利用した Morphing により高速な自由視点画像生成を行う．以下にその処理の概要を示す．なお，全方位カメラの位置・姿勢および入力画像間の対応点はあらかじめ与えられているものとする．

Step1. 与えられた各対応点について，それらの点の 3 次元位置を全方位ステレオにより計算する．

Step2. Step1 で得られた 3 次元点を式 (2) の計算により自由視点画像上に投影する ( 図 42 参照 ) ．

Step3. 投影された自由視点画像上の点群に対してドロネーの三角分割法 [Hec94] により三角形パッチを生成する．

Step4. Step3 で得た三角形パッチに対応する三角形の画像を各入力画像から切り出し，各実カメラから自由視点位置へのベクトルの重み付け線形和が自由視点位置となるような重みを用いてブレンドし，自由視点位置における全方位画像とする．三角形パッチ内部の画素に対するブレンディング処理については，OpenGL を用いて GPU による高速な計算を行う．

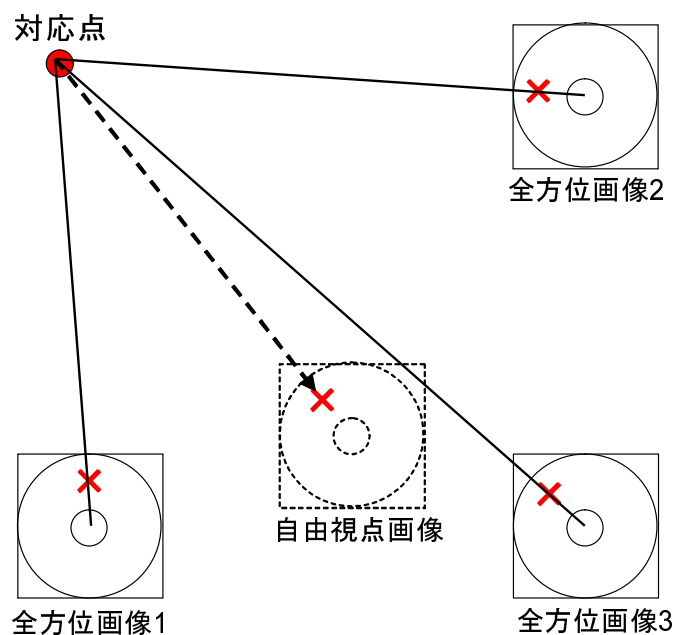


図 42 対応点の自由視点画像への投影

### 3.4.4 動的領域に対する Visual Hull 推定による自由視点画像生成

3.4.1 節で生成された動的領域に対して，Visual Hull を用いた自由視点画像生成を行う．Visual Hull とは，多視点から撮影された画像群  $R$  中の物体のシルエット  $s_r$  とカメラの光学中心によって形成される錐体  $vh_r$  の積集合空間  $vh_R = \cap_{r \in R} vh_r$  を指す．一般に，Visual Hull は空間を立方体で区切ったボクセルを用いて表現されることが多い[ウ 01, 上田 05]．しかし，ボクセルを用いた表現では，本研究で対象とするように撮影対象の空間が広がるとそのデータ量と計算量が膨大となる．そこで本研究では，Visual Hull を表現する手法に Image-based Visual Hull[MBR<sup>+</sup>00]を用いる．この手法は，自由視点における画像面の各画素について，その画素と自由視点位置とを結ぶ空間直線が，Visual Hull を貫通するか否かを判定することにより自由視点画像を生成する．以下に Image-based Visual Hull[MBR<sup>+</sup>00] の処理の概要を示す．本研究では，全方位画像を用いて Visual Hull を計算しているが，ここでは説明を簡単にするために通常のカメラを用いた場合について述べる．

Step1. 自由視点画像面のある画素について，その画素と自由視点を結ぶ空間直

線を各実カメラの画像面上に投影する．各実カメラの画像面上に投影された自由視点がエピポール，投影直線がエピポーラ線となる．

Step2. 各実カメラに投影されたエピポーラ線上で動的領域と交差する線分を調べ，すべてのエピポーラ線で動的領域と交差する線分があるならば，自由視点からの空間直線は Visual Hull を貫通する．交差線分がないならば空間直線は Visual Hull を貫通しないと判断できる（図 43 参照）．

Step3. 自由視点からの直線が Visual Hull を貫通する画素については，その画素の色を決定するために，交差線分の最も自由視点位置に近い点への各実カメラからのベクトルと自由視点からのベクトルとを比較し，それらのベクトルの成す角が最も小さい実カメラの画素の色を処理対象としている画素の色とする．

Step4. Step1～3 の処理を自由視点画像面上のすべての画素に対して行う．

以上の処理により動的領域の全方位の自由視点画像を生成可能である．しかし，自由視点画像上の全画素について Image-based Visual Hull[MBR<sup>+</sup>00] の計算を行うと計算量が多く，実時間での画像生成は困難となる．そこで計算量削減のために，3.4.2 節で推定した動的領域の存在領域を基に Visual Hull の計算が必要な自由視点画像上の画素を限定する．存在領域推定処理によって計算された直線の交点を式 (2) を用いて自由視点画像面上に投影し，投影された交点が存在する方位角範囲の画素についてのみ Visual Hull の計算を行う．さらに，自由視点からの空間直線が Visual Hull を貫通するか否かを評価する際，エピポーラ線上の推定された動的領域の存在領域以外では Visual Hull の評価を省略する（図 44 参照）．

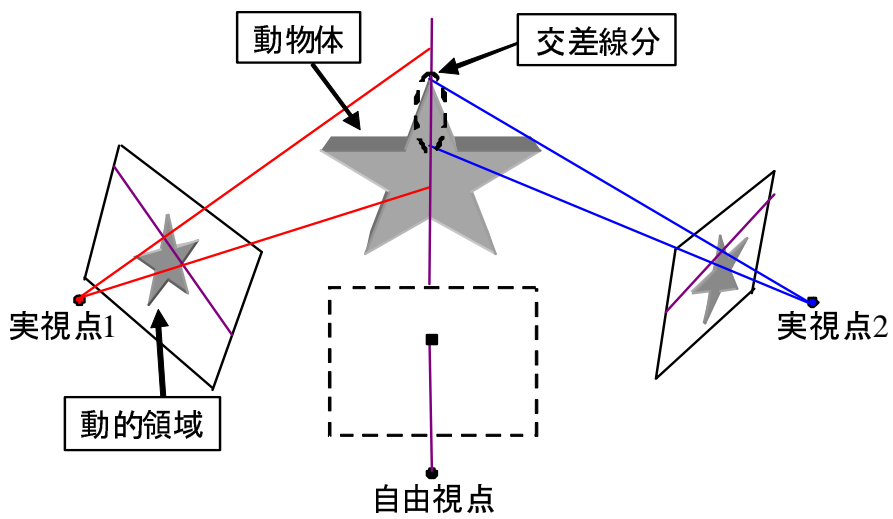


図 43 空間直線の Visual Hull の貫通判定

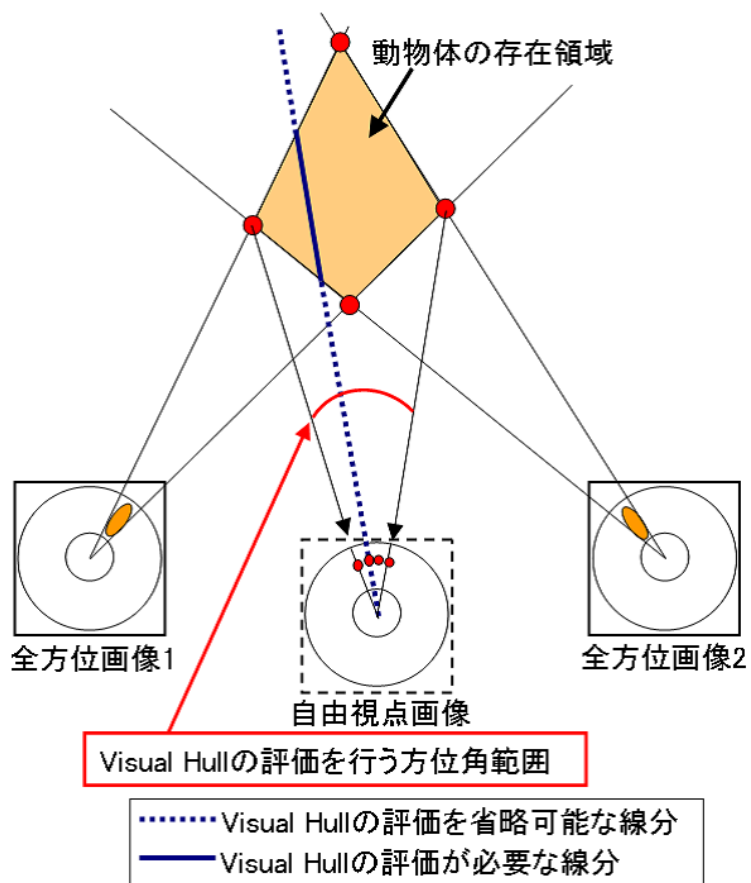


図 44 動物体の存在領域推定による処理の削減

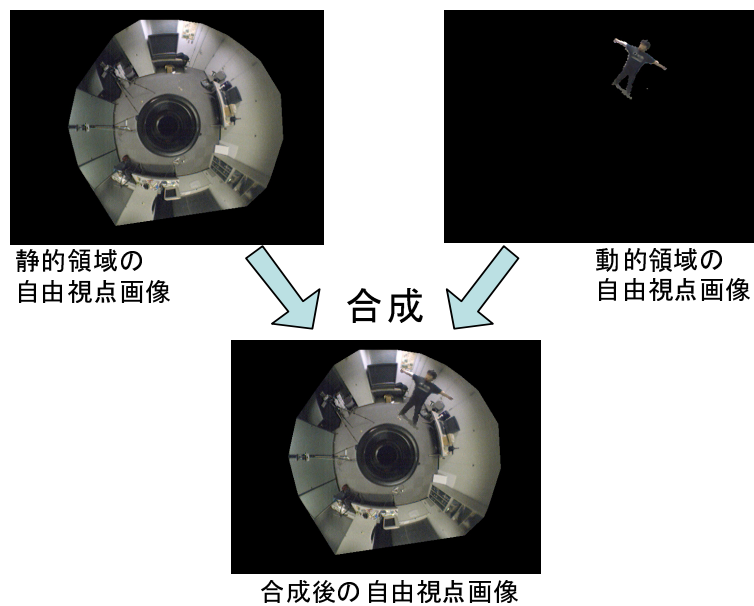


図 45 静的領域と動的領域の自由視点画像の合成

### 3.4.5 静的領域と動的領域の自由視点画像の統合

3.4.3, 3.4.4 節の処理により, 自由視点位置における静的領域と動的領域の全方位画像が得られる. この二つの自由視点画像を合成し, 最終的な自由視点画像とする (図 45 参照). 合成処理は静的領域の自由視点画像の上に動的領域の自由視点画像を重ね合わせる (上書きする) 簡単な処理である. ただし, 動的領域の自由視点画像上の Visual Hull が投影されなかった画素は合成処理を行わない.

### 3.5 シミュレーションによる定量的画質評価に基づくカメラ配置の検討

本節では提案する自由視点画像生成手法において、全方位カメラの配置や撮影環境、カメラの解像度等の要因が生成する自由視点画像の画質に与える影響を、シミュレーション環境を用いた定量的画質評価により検証する。そして、それらの結果から想定する撮影環境に対して、システムの利用者の要求を満たすためにはどのようにカメラを配置すべきかを考察する。また検証結果に基づき、撮影環境と自由視点位置の移動可能範囲の関係から、カメラ配置の定量的評価指標を示す。

提案した自由視点画像生成手法は、全方向に物体が存在する観光地のような環境での使用を想定している。シミュレーションではこのような環境を3DCGオブジェクトにより構築し、この環境中で描画した全方位画像を入力としてある自由視点位置における自由視点全方位画像を生成する。この自由視点全方位画像とその自由視点位置で描画した全方位画像とでPSNR(Peak Signal-to-Noise Ratio)による定量的画質評価を行う。以下に、定量的画質評価実験の目的と条件、画質評価結果とその結果に基づくカメラ配置の検討について述べる。

#### 3.5.1 実験目的と条件

提案手法により生成される自由視点画像の画質に影響を与える要因とその影響を明らかにすることは、システム的设计や撮影環境に対するカメラ配置の決定を行う上で重要である。本実験では、画質に影響を与える要因とそれらの要因による影響を明らかにすることを目的とする。そして、それらの結果に基づきシステムの利用者が撮影環境にカメラを配置する際、どのようにカメラを配置すべきかを検討するための指標を示す。

提案手法により生成される自由視点画像の画質に影響を与える要因として以下のものが考えられる。

- 撮影環境の形状



- 撮影環境のテクスチャ
- 撮影対象とカメラ群との距離
- カメラ配置（各カメラの位置・カメラ間の距離）
- カメラの解像度
- カメラの数
- 自由視点位置

本実験では、定量的な画質評価指標である PSNR により上記の要因が自由視点画像の画質に与える影響を検証する。しかし、上記要因のうち撮影環境の形状とテクスチャについてはそれらを定量的に評価することが困難であるため、シミュレーションでは提案した自由視点画像生成手法の使用が想定されるような環境を構築し実験を行う。3DCG オブジェクトにより構築したシミュレーション環境を図 46 に示す。この環境中で、その他の要因のパラメータを以下のように変化させ画質の評価を行う。

(A) 撮影対象とカメラ群との距離

動物体の位置をカメラ群の重心位置から、約 0.65, 1.3, 2.6m の距離に配置する。

(B) カメラ配置（各カメラの位置・カメラ間の距離）

カメラ群の重心位置を固定してカメラ間の距離を、1, 2, 4m とする。

(C) カメラの解像度

512 × 512, 768 × 768, 1024 × 1024 画素とする。

(D) カメラの数

3 台または 4 台とする。

(E) 自由視点位置

カメラ群から成る三角形内の格子点位置とする。

ただし、各要因を変化させる際には基準となる各要因のパラメータパターン（以下、基準パターン）から各要因のパラメータを変化させる。基準パターンとして各要因のパラメータを以下のように設定する。

- (A) 撮影対象とカメラ群との距離：約 1.3m
- (B) カメラ配置（各カメラの位置・カメラ間の距離）：2m
- (C) カメラの解像度：512 × 512 画素
- (D) カメラの数：3 台
- (E) 自由視点位置：カメラ群から成る三角形内の格子点位置

基準パターンにおける撮影対象・カメラ位置・自由視点位置の関係を図 47 に示す。カメラは、図 47 中赤丸で示した点に配置しており、その座標は環境中の地面と平行な面を XY 座標面、高さ方向を Z 軸としている。また、本実験でカメラ群は正三角形に配置する。これはあるカメラ間の距離に対して自由視点位置が移動可能な面積が最大となるカメラ配置であることや、カメラ間の距離の差が画質へ与える影響を排除するためである。自由視点位置については基準パターンに限らず、X 座標、Y 座標それぞれについてカメラ間の距離を 20 等分したときの格子点位置とする（図 47 中の青点）。また、動物体として手を広げた人物を配置した。

各要因のパラメータに関わらず共通のパラメータとして、ミラーの形状パラメータ  $a, b, c$  をそれぞれ 26.26mm, 35.49mm, 44.15mm とする。これらの値は、3.6 節の実験で用いる全方位カメラのパラメータとほぼ同等である。全方位カメラの焦点距離は、カメラの画角が解像度によって変化しないよう設定する。また、各入力全方位画像の解像度と出力自由視点全方位画像の解像度は同じとする。図 48 に、基準パターンにおける入力全方位画像と静的領域の対応点を示す。図 48 上段は静的領域のみが含まれる全方位画像であり、その図中の × 印は対応点を示している。ここでは手動で約 160 点の対応点を与えている。ただし、静的領域の対応点はカメラ配置やカメラ数が異なる場合には、その都度新たに対応点を与えている。図 48 下段は動的領域のみが含まれる全方位画像である。

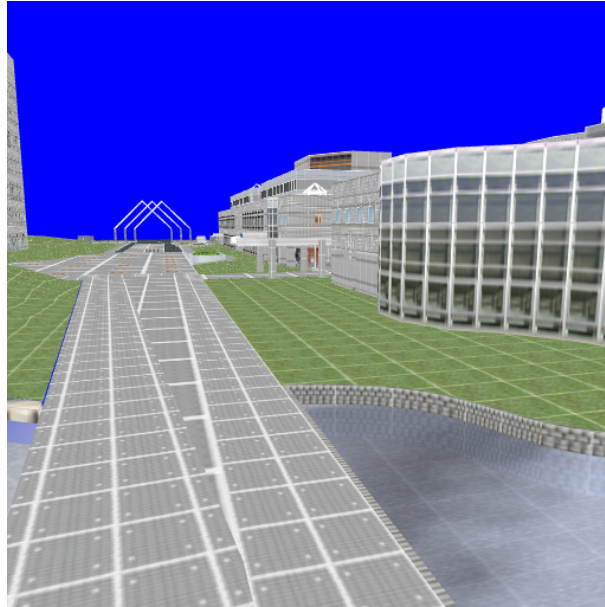


図 46 観光地を想定したシミュレーション環境

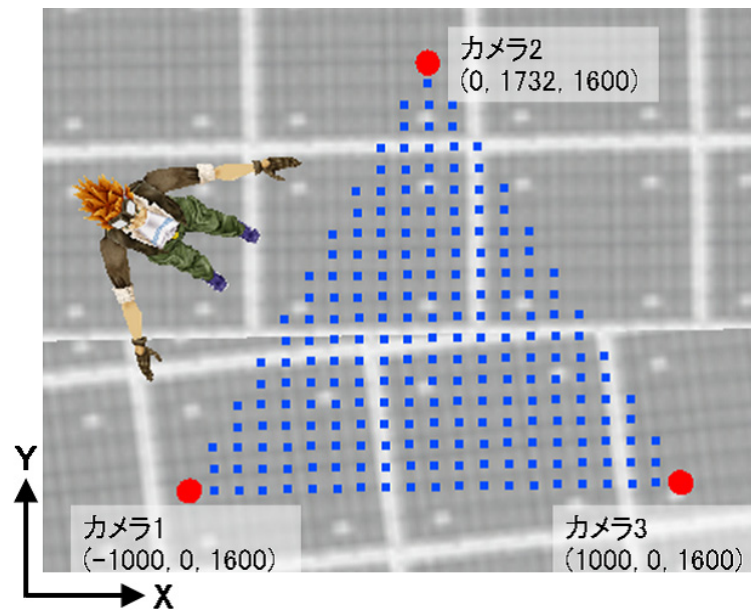


図 47 シミュレーション環境中の撮影対象（人物）とカメラ配置（赤丸）及び自由視点位置（青点）

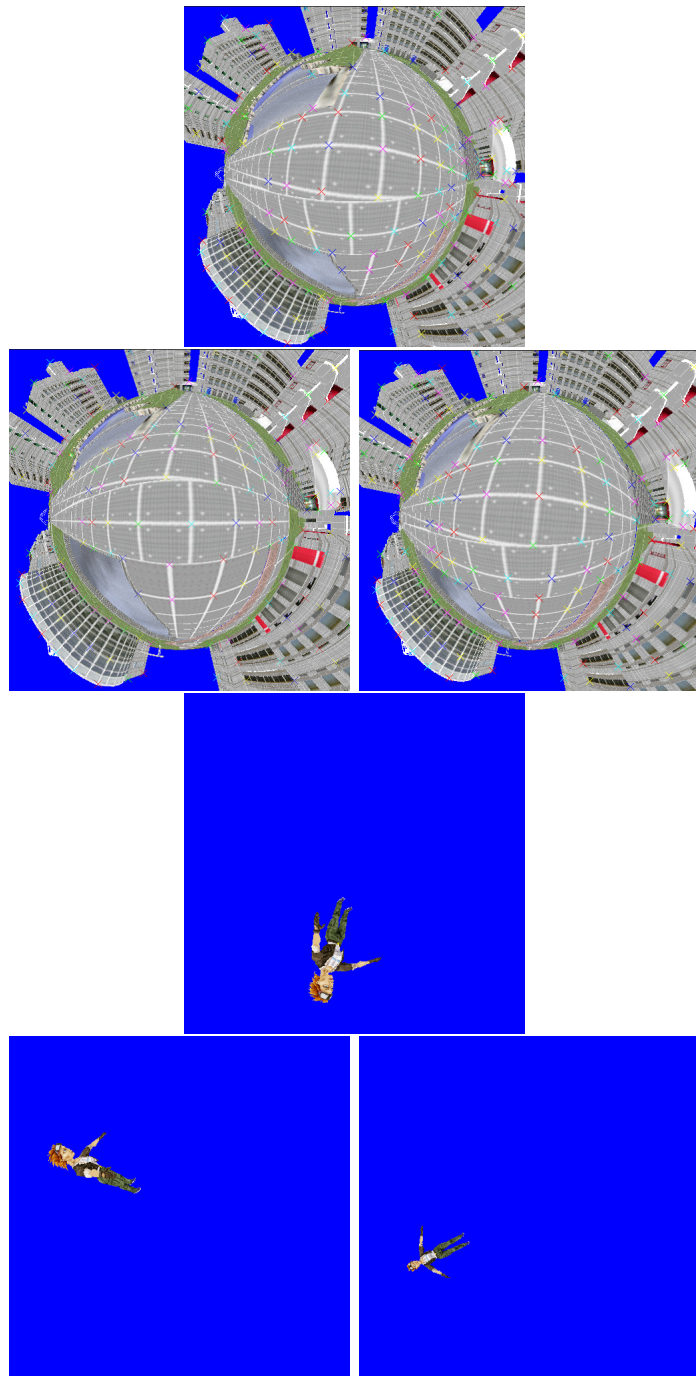


図 48 基準パターンにおける静的領域の画像とその対応点（上段）および動的領域の全方位画像（下段）

本実験では，定量的画質評価の指標として先に述べたように PSNR を用いる．PSNR は以下の式で定義され，自由視点全方位画像とその真値となる全方位画像の一致度を評価する．

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right) \quad (9)$$

MAX は画像の最大輝度値（各チャンネル 8bit の画像の場合 255），MSE（Mean Squared Error）は自由視点全方位画像と真値の全方位画像の平均二乗誤差である．PSNR が大きな値となるほど両者に差が少なく，生成された自由視点全方位画像は高画質と言える．以下では，基準パターンから各要因を変化させたときの PSNR による評価結果を示す．そして，その評価結果を考察する．

### 3.5.2 定量的画質評価結果とカメラ配置に関する考察

まず，基準パターンにおける画質評価結果を示す．図 49 に，各自由視点位置における PSNR の値をカラーマップとして示す．図中の色はその位置で生成された自由視点画像の画質を表し，暖色系の色であるほど画質が高く，寒色系の色であるほど画質が低いことを表す．このとき PSNR の最大値は 47.29[dB]，最小値は 20.36[dB]，平均値は 22.14[dB] であった．図 49 から，自由視点位置が入力画像が撮影された位置に近いほど高品質の画像が生成できていることがわかる．特に，自由視点位置が配置されたカメラの位置にある場合には，生成された自由視点画像と真値の画像の差を視認することはできなかった．生成画像の画質が低い位置は，動物体の付近に集中している．これは，自由視点画像上の動物体の表現に Visual Hull を使用しているため，動物体の本来の形状と Visual Hull による形状の差が原因となっている．この形状の差は，自由視点位置が動物体に近づくにつれて画像上に大きく現れるため，PSNR が低くなっている．しかし，図 47 に示すカメラ 3 と動物体を結ぶ直線上の自由視点位置では，カメラ 3 からの入力画像と自由視点画像との視差が小さく，形状の差が少なくなるためその他の視点位置に比べ若干画質が良い結果となっている．図 50 に生成した自由視点全方位画像（左）とそれに対応する真値の全方位画像（右）の例を示す．以下では，要因（A）～（E）のパラメータを変化させた場合の結果について述べる．

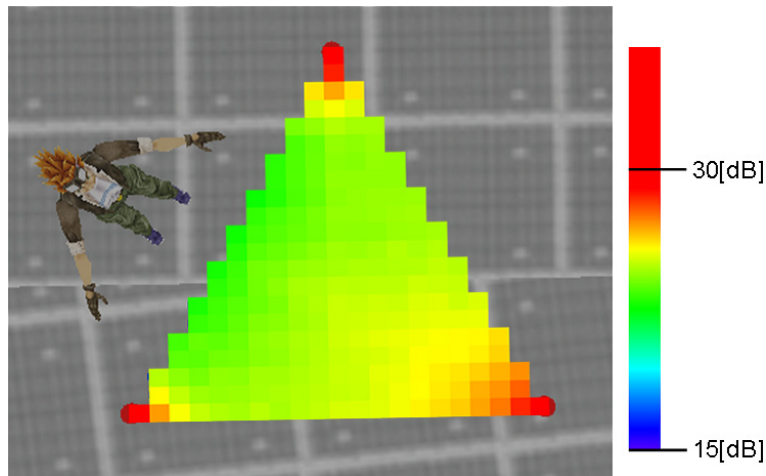
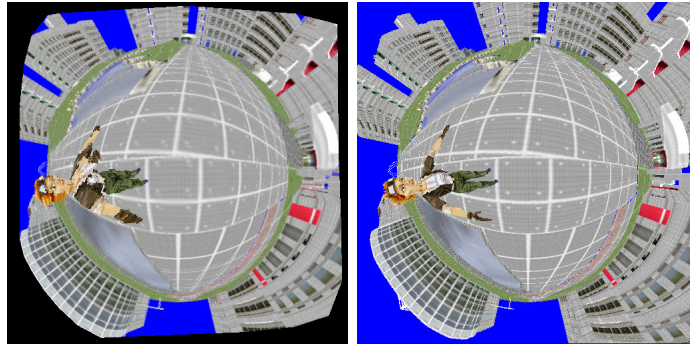


図 49 基準パターンにおける PSNR カラーマップ

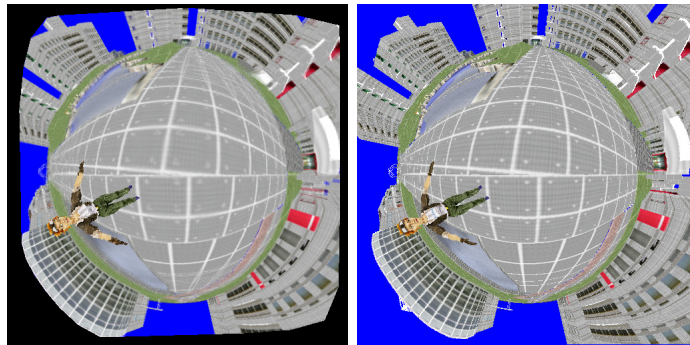
[要因 (A): 撮影対象とカメラ群との距離を変化させた場合]

動物体位置とカメラ群の重心位置との距離を約 0.65, 1.3, 2.6m と変化させたとき, 生成した自由視点画像の画質がどのように変化するかを検証する. 図 51 に, 動物体とカメラの位置関係を示す. 動物体位置 0 のとき, 動物体とカメラ群の重心位置との距離は約 1.3m であり, 基準パターンと同じ環境となる. 動物体位置 1, 2 のとき, 動物体とカメラ群の重心位置との距離はそれぞれ約 0.65, 2.6m である.

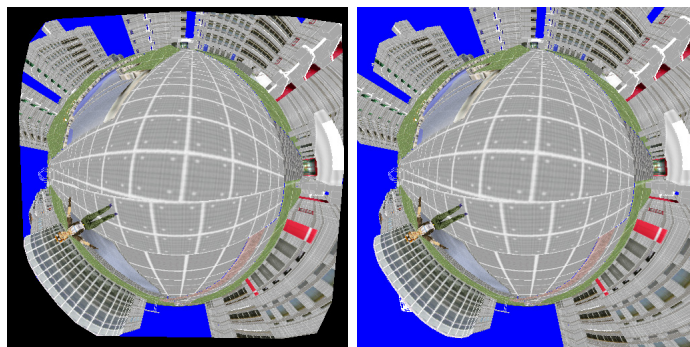
要因 (A) の各パラメータにおける画質評価結果を図 52 に示す. 図 52 上段が動物体位置 1, 中段が基準パターンと同様の動物体位置 0, 下段が動物体位置 2 における PSNR のカラーマップと平均・最大・最小値である. 動物体がカメラ群の重心位置から離れるに従い画質が上がっていることが確認できる. また, 基準パターンでの画質評価の場合と同様にカメラ位置から自由視点位置が離れると画質が低下している. 動物体とカメラ群重心との距離が約 0.65m のとき, PSNR の最小値が 16.19[dB] となっており真値と大きく異なる画像となっていた. これは, 動物体の真の形状と Visual Hull の形状との差が自由視点画像上に大きく投影されるためである. 動物体がカメラ群から離れると動物体の形状と Visual Hull の形状の差があった場合でも, その差は自由視点画像上では小さく投影されるため画質に与える影響は少ない.



(-300, 1039, 1600) : PSNR 20.36[dB]



(0, 606, 1600) : PSNR 21.78[dB]



(1000, 0, 1600) : PSNR 47.29[dB]

図 50 生成した自由視点全方位画像（左）と真値の全方位画像（右）

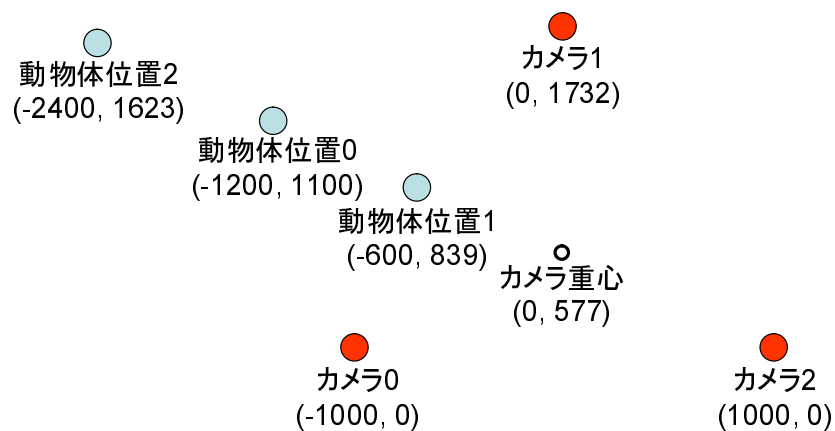


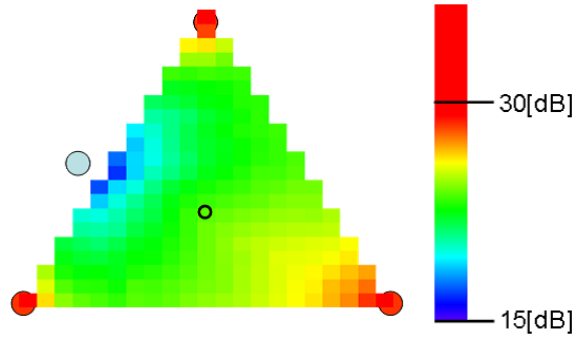
図 51 撮影対象（動物体）とカメラ間の距離を変化させた場合の動物体・カメラの位置関係

[要因 (B): カメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合]

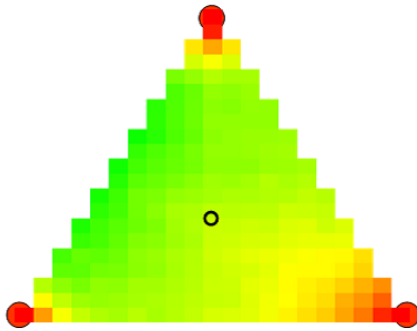
カメラ群の重心位置を固定してカメラ間の距離を，1，2，4m と変化させたとき生成される自由視点画像の画質がどのように変化するかを調べる．図 53 に，カメラ配置を変化させた場合のカメラの位置関係を示す．図中赤丸で示すカメラ配置はカメラ間の距離が 2m であり，基準パターンと同様である．図中緑丸はカメラ間の距離を 1m とした場合のカメラ配置，青丸はカメラ間の距離を 4m とした場合のカメラ配置を示す．また，環境中の動物体の位置は全てのカメラ配置において基準パターンと同様である．

要因 (B) の各パラメータにおける画質評価結果を図 54 に示す．図 54 上段がカメラ間距離が 1m，中段が基準パターンと同様のカメラ間距離 2m，下段がカメラ間距離が 4m の場合の PSNR のカラーマップと平均・最大・最小値である．カメラ間距離が短いほど画質が高くなる傾向が確認できる．これは要因 (A) を変化させた場合と同様に，カメラ間距離が短くなることで自由視点画像と入力画像間の視差が小さくなるからだと考えられる．カメラ間を 4m とした場合，PSNR の最小値が 10.17[dB] となっており大きな画質の劣化が確認できる．これは自由視点位置と動物体の距離が非常に近くなり，要因 (A) の結果と同様に大きく画質が劣化するためである．また，動物体が含まれない場合の画質評価も実施した

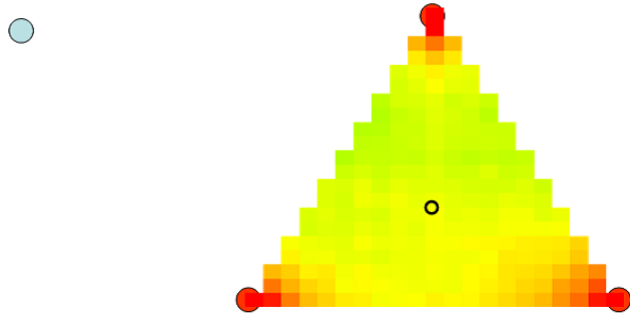




PSNR : 平均値 21.03[dB] , 最大値 46.46[dB] , 最小値 16.19[dB]



PSNR : 平均値 22.14[dB] , 最大値 47.29[dB] , 最小値 20.36[dB]



PSNR : 平均値 22.76[dB] , 最大値 46.30[dB] , 最小値 21.84[dB]

図 52 撮影対象（動物体）とカメラ群重心との距離を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：約 0.65m，中段：約 1.3m，下段：約 2.6m）

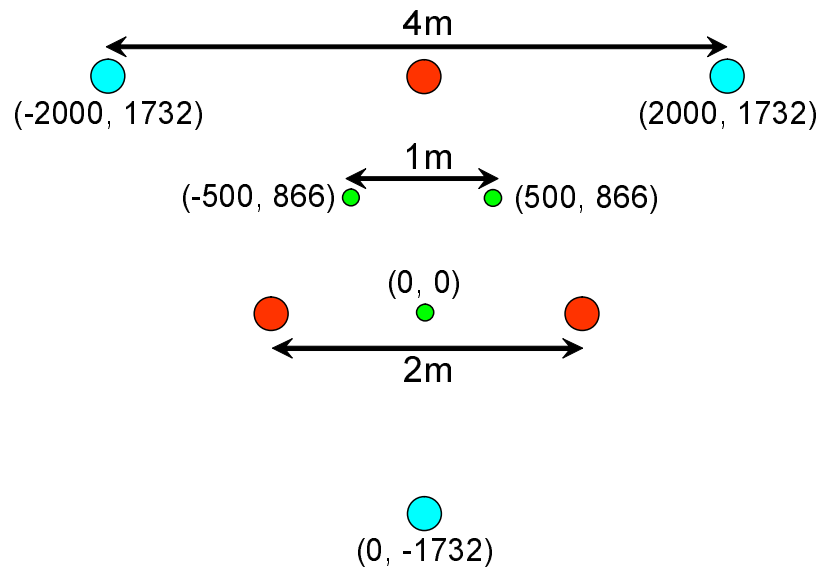


図 53 カメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合のカメラの位置関係（赤丸：基準パターンと同じカメラ配置，緑丸：カメラ間距離が 1 m のカメラ配置，青丸：カメラ間距離が 4 m のカメラ配置）

（図 55 参照）．動物体が含まれる場合同様に，カメラ間の距離が短いほど画質が高くなる傾向が確認できた．

[要因 (C) : カメラの解像度を变化させた場合]

カメラの解像度を  $512 \times 512$  ,  $768 \times 768$  ,  $1024 \times 1024$  画素に変化させたとき生成される自由視点画像の画質がどのように変化するかを調べる．動物体とカメラの位置関係は基準パターンと同様である．また，自由視点画像の解像度は入力全方位画像の解像度を同じとする．

要因 (C) の各パラメータにおける画質評価結果を図 56 に示す．図 56 上段が解像度を  $512 \times 512$  画素，中段が  $768 \times 768$  画素，下段が  $1024 \times 1024$  画素の場合の PSNR のカラーマップと平均・最大・最小値である．カメラの解像度を高くするに従い画質が向上する傾向にあることが分かる．これは図 57 に示すように，自由視点画像と真値の全方位画像間で誤差が生じる領域が自由視点画像上のエッジ周辺であるのに対して，解像度が高くなることでエッジ以外の領域の割合が増加し，

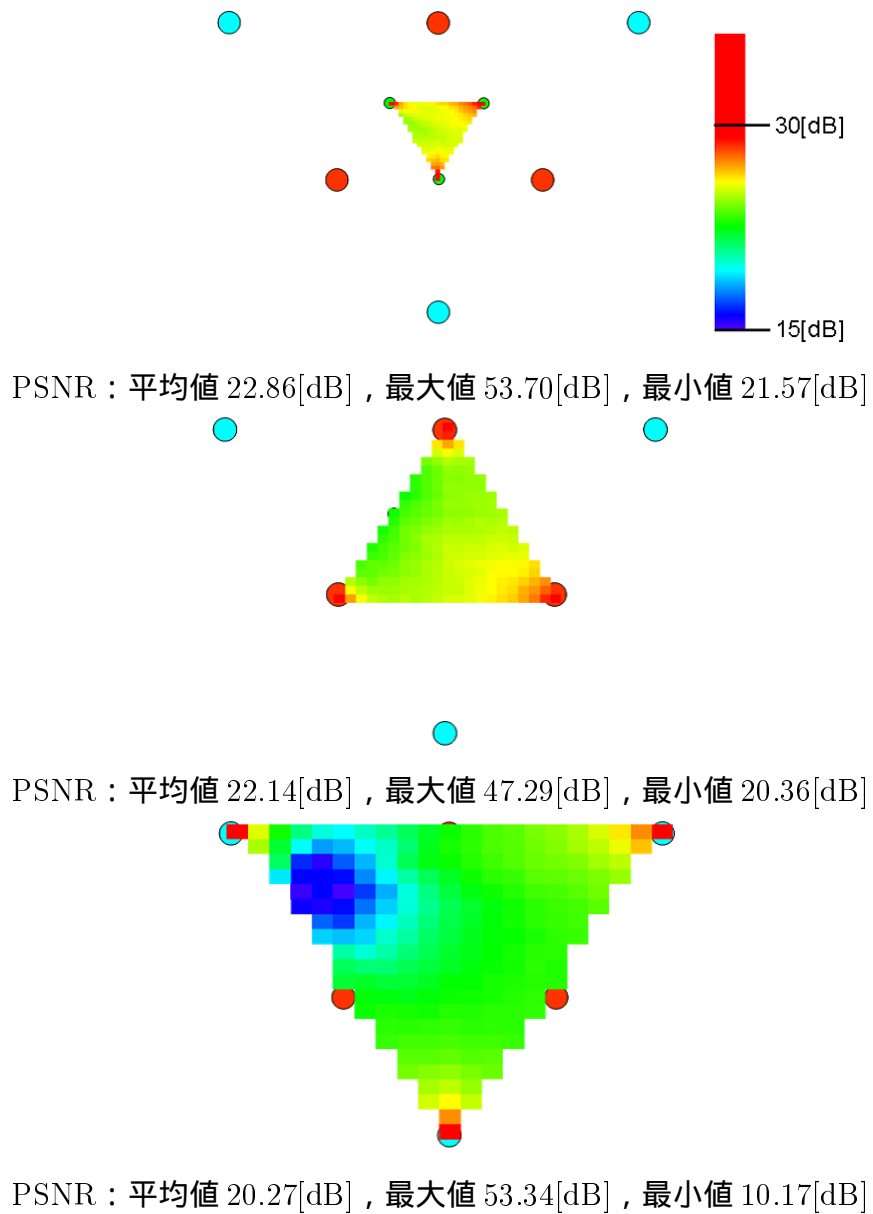


図 54 カメラ配置 (各カメラの位置・カメラ間の距離) を変化した場合の PSNR のカラーマップと平均・最大・最小値 (上段 : カメラ間距離 1m , 中段 : カメラ間距離 2m , 下段 : カメラ間距離 4m )

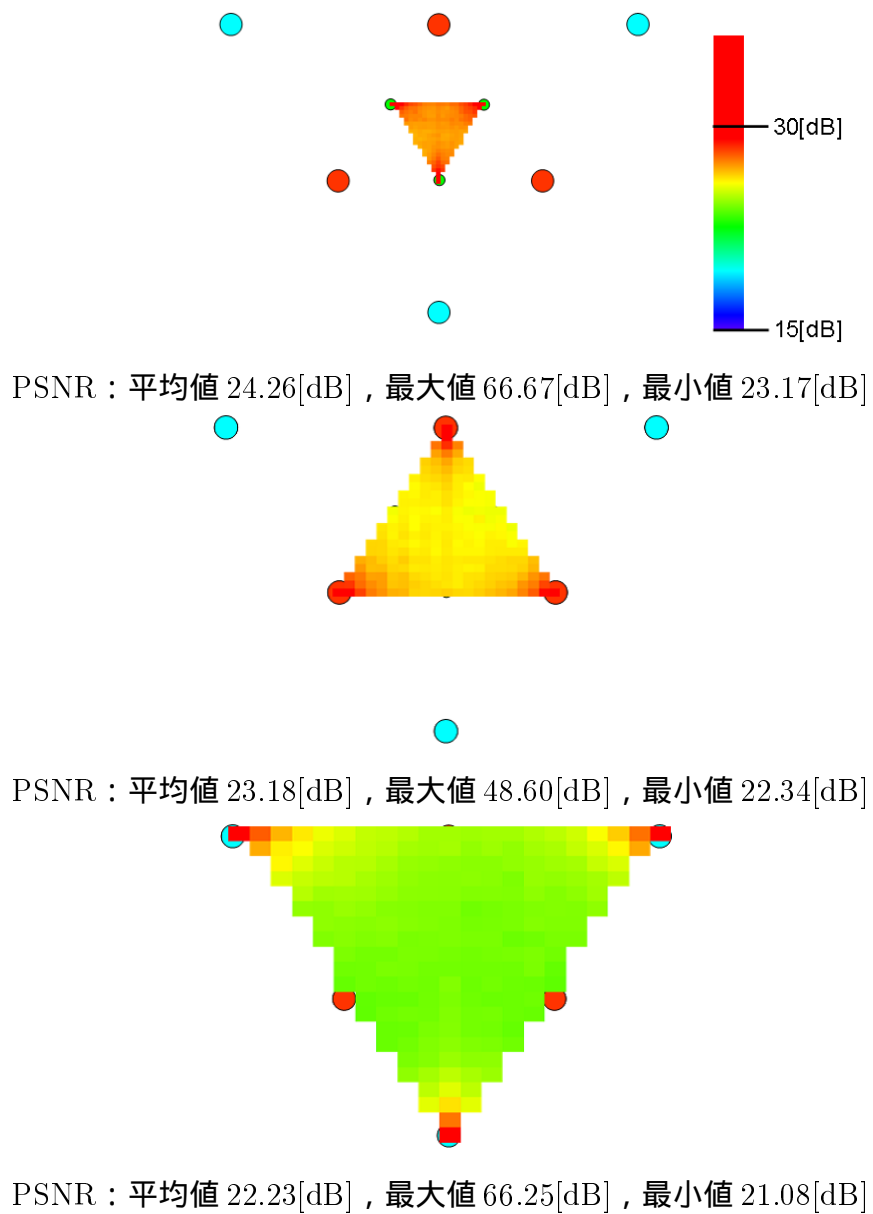


図 55 動物体を含まずカメラ配置（各カメラの位置・カメラ間の距離）を変化させた場合の PSNR のカラーマップと平均・最大・最小値（上段：カメラ間距離 1m，中段：カメラ間距離 2m，下段：カメラ間距離 4m）

画像全体での画質が向上したと考えられる．一方，解像度を高くすることで計算コストが増加する．特に自由視点画像生成処理において Image-based Visual Hull は計算コストの大部分を占め，解像度に対して線形に計算コストが増加する．

[要因 (D): カメラの数を変化させた場合]

自由視点画像生成に用いるカメラの数を 3 台から 4 台に変化させたとき生成される自由視点画像の画質がどのように変化するかを検証する．要因 (D) の各パラメータにおける画質評価結果を図 58 に示す．図 58 上段がカメラ数が 3 台，下段が 4 台の場合の PSNR のカラーマップと平均・最大・最小値である．図 58 下段中の青丸が基準パターンに対して追加したカメラ位置を示している．カメラ数を増加させた場合，若干の画質が向上が確認できるもののほとんど変化がなかった．この原因としては，動物体を撮影する視点数が増加したことで Visual Hull がより真の形状に近づくが，自由視点位置と追加されたカメラ間の距離が離れている場合には，Morphing 処理 (3.4.3 節) において追加したカメラの重みが小さくなるため，3 台のカメラを使用した場合と比較して画質の差が生じにくかったことが考えられる．また，カメラ数が増加することで自由視点画像生成に必要な計算コストが増加する．カメラ台数が 3 台から 4 台に増加した場合，計算コストの増加は約 33 % である．

[画質評価結果に関する考察]

以上の画質評価結果から，次のことが確認できた．

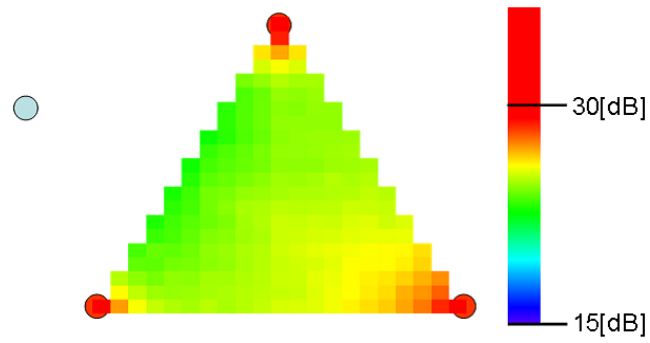
(A) 撮影対象とカメラ群との距離

撮影対象 (動物体) がカメラ群から離れた位置に存在する場合，撮影対象の真の形状と Visual Hull により推定された形状の差が自由視点画像上で顕著となり難いため，画質に与える影響は小さくなる．

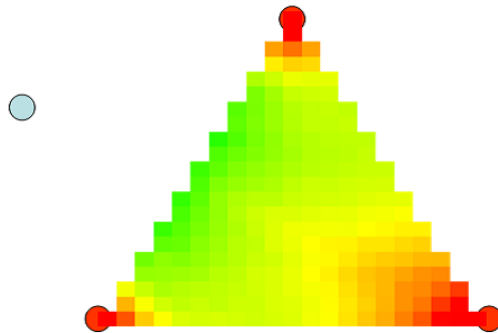
(B) カメラ配置 (各カメラの位置・カメラ間の距離)

カメラ間の距離を短くするほど，入力全方位画像と自由視点全方位画像間の視差が小さくなるため，画質が向上する．

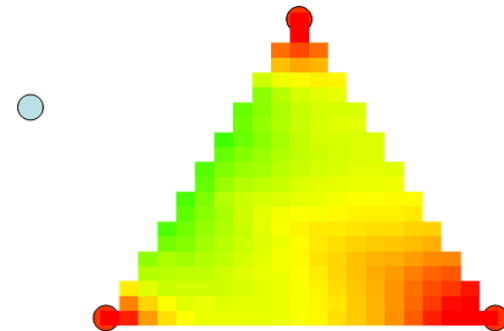
(C) カメラの解像度



PSNR : 平均値 22.14[dB] , 最大値 47.29[dB] , 最小値 20.36[dB]



PSNR : 平均値 22.54[dB] , 最大値 46.49[dB] , 最小値 20.61[dB]



PSNR : 平均値 22.79[dB] , 最大値 47.34[dB] , 最小値 20.82[dB]

図 56 カメラの解像度を変化させた場合の PSNR のカラーマップと平均・最大・最小値 ( 上段 : 512 × 512 画素 , 中段 : 768 × 768 画素 , 下段 : 1024 × 1024 画素 )

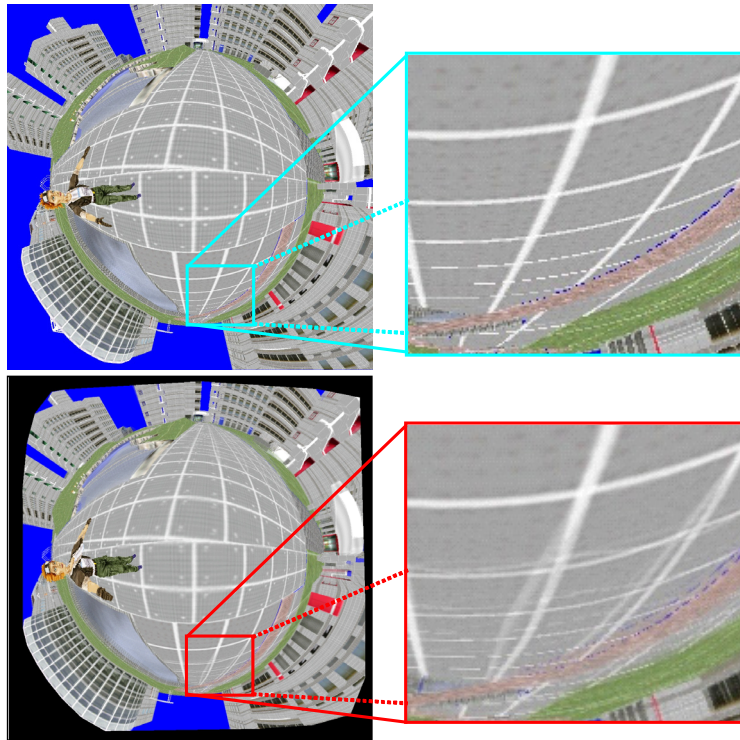


図 57 エッジ周辺での画質劣化例（上段：真値の全方位画像，下段：自由視点全方位画像）

カメラの解像度を高くすることで、誤差の生じやすいエッジ領域に比べそれ以外領域の割合が増加するため画質が向上する。ただし、計算コストは解像度に対して線形に増加する。

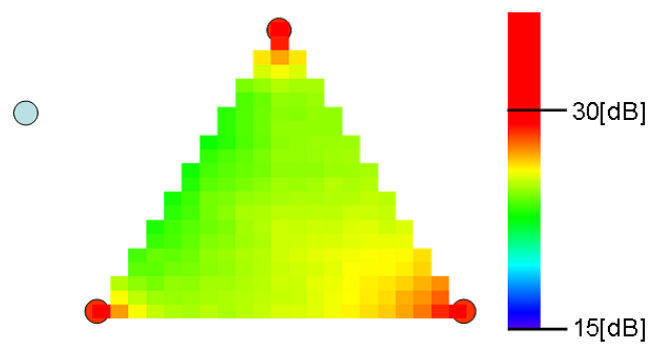
（D）カメラの数

カメラ数を増加させることで、若干の画質向上があるものの、ほとんど変化がない。また、計算コストはカメラ台数の増加に従い増加する。

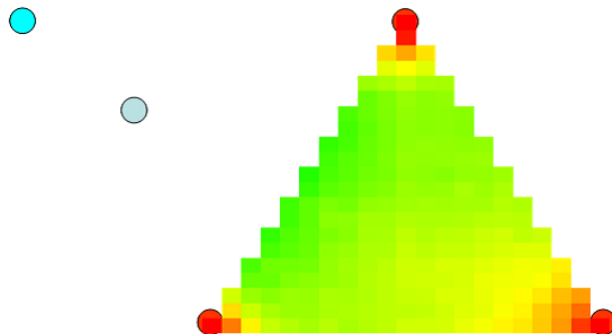
（E）自由視点位置

自由視点位置がカメラに近いほど、入力全方位画像と自由視点全方位画像間の視差が小さいため画質の高い自由視点画像が生成される。

これらの各要因と画質の関係から、システムの利用者は環境にカメラを配置する際に、あるカメラ配置に対してより画質を高くする必要がある場合には、撮影対



PSNR : 平均値 22.14[dB] , 最大値 47.29[dB] , 最小値 20.36[dB]



PSNR : 平均値 22.20[dB] , 最大値 51.34[dB] , 最小値 20.55[dB]

図 58 カメラの数を変化させた場合の PSNR のカラーマップと平均・最大・最小値 (上段 : 3 台 , 下段 : 4 台)



象とカメラ群の距離を離すこと，カメラ間の距離を短くすること，カメラの解像度を高くすることで自由視点画像の画質を高めることが可能である．ただし，撮影対象とカメラ群を離すことで入力画像と自由視点画像間の視差が小さくなるため，画像を観賞する際に視点を移動している感覚が減少することが考えられる．また，カメラ間の距離を短くすることで同じカメラ数では視点移動範囲が狭くなるため，視点移動範囲を保つためにはカメラの台数を増やす必要がある．カメラの解像度を高めることについては計算コストの増加に繋がるため，インタラクティブな映像提示を行うためには十分に高性能な計算機を利用する必要がある．

システムの利用者がより広い範囲を視点移動可能にしたい場合には，カメラ間の距離を離すこと，カメラ数を増やすことで対応可能である．ただし，カメラ間の距離を離すことで画質が低下するため，画質を保持するためには撮影対象とカメラ群の距離を長くすることやカメラの解像度を高くすることによる対応が必要である．また，カメラ数を増やすことによって，計算コストが増加するためそれに対応可能な高性能な計算機を利用する必要がある．自由視点画像の画質や視点移動可能範囲の要求に対しては上記のように対応することができる．

撮影環境によっては利用者が想定するようなカメラ配置が困難であることが考えられる．このような場合，本実験のようにカメラ群を正三角形に配置できるとは限らない．しかし，カメラ群による三角形がどのような形状であっても，上に挙げた各要因と画質の関係から画質の変化が予測可能である．例えば，想定したカメラ配置が正三角形であったが，カメラ設置の都合上視点移動可能な面積を縮小し，カメラ配置を変更する必要がある場合，想定していたカメラ配置に比べカメラ間の距離が短くなるため画質が向上すると考えられる．

各要因について考察すると，要因(A)・(B)・(E)は入力画像と自由視点画像間の視差に関するパラメータと言える．要因(C)は撮影環境のテクスチャに依存するパラメータ，要因(D)はVisual Hullの形状に関するパラメータと言える．画質評価結果から，要因(A)・(B)・(E)によるPSNRの変動は0.62～43.17[dB]，要因(C)による変動は0.4[dB]以下，要因(D)による変動は0.06[dB]となっており，入力画像と自由視点画像間の視差が最も画質に影響を与える要因であることが分かる．そして，その視差が大きいほど画質が低下し，小さいほど画質が高

くなる傾向が確認できる．そこで本研究では，撮影環境に対してどのようにカメラを配置するかを決定するための入力画像と自由視点画像間の視差に基づく定量的評価指標を提案する．次節に，提案するカメラ配置の定量的評価指標とその評価指標による最適なカメラ配置の推定について述べる．

### 3.5.3 視差に基づくカメラ配置の定量的評価指標と最適なカメラ配置の推定

前節での画質評価結果から入力画像と自由視点画像間の視差が画質に最も影響を与える要因であることが分かり，視差が小さいほど画質が高くなる傾向となることが分かった．本節では，入力画像と自由視点画像間の視差に基づき撮影環境に対して最適なカメラ配置を決定するための評価指標について述べる．そして，その評価指標による最適なカメラ配置の推定方法と模擬環境における推定例について述べる．

[視差に基づくカメラ配置の定量的評価指標]

ある撮影環境  $G$  に対して設定された自由視点の移動可能範囲  $V$  において，自由視点画像  $\{W_v | v \in V\}$  の画質を最も高くするカメラ群  $C$  の配置  $\{P_c | c \in C\}$  は，以下の評価指標  $E_{parallax}$  の非線形最小化により得られる．

$$E_{parallax} = \frac{1}{||V||} \sum_{v \in V} \left( \frac{1}{||W_v||} \sum_{s \in W_v} \sum_{c \in C} \omega_c D(G_s, P_c, P_v) \right) \quad (10)$$

ここで  $s = \{u, v\}$  は画像上の座標， $G_s = \{X_s, Y_s, Z_s\}$  は  $s$  に投影される環境中の三次元点， $P = \{X, Y, Z\}$  は3次元空間中の座標を示す．そして， $\omega_c$  は3.4.3節で述べた各カメラへの重み係数， $D(G_s, P_c, P_v)$  は視差を計算する関数である． $\omega_c$  と  $D(G_s, P_c, P_v)$  はそれぞれ以下の式で与えられる．

$$\begin{bmatrix} X_0 & X_1 & X_2 \\ Y_0 & Y_1 & Y_2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{bmatrix} = \begin{bmatrix} X_v \\ Y_v \\ 1 \end{bmatrix}$$

$$\mathbf{A} \begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{bmatrix} = \begin{bmatrix} X_v \\ Y_v \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} X_v \\ Y_v \\ 1 \end{bmatrix} \quad (11)$$

ただし,  $\|C\| = 3$

$$D(\mathbf{G}_S, \mathbf{P}_c, \mathbf{P}_v) = \arccos \left( \frac{(\mathbf{P}_c - \mathbf{G}_S)(\mathbf{P}_v - \mathbf{G}_S)}{\|(\mathbf{P}_c - \mathbf{G}_S)\| \|(\mathbf{P}_v - \mathbf{G}_S)\|} \right) \quad (12)$$

$E_{parallax}$  は, 自由視点画像上の視差の平均を視点移動可能範囲内の自由視点数で平均化した評価指標であり, これが小さいほど視差が小さくなり画質の高い自由視点画像が得られる. 次に,  $E_{parallax}$  の非線形最小化処理について述べる.

[ $E_{parallax}$  の最小化による最適なカメラ配置の推定]

撮影環境の形状  $G$  と自由視点の移動可能範囲  $V$  が与えられたとき,  $E_{parallax}$  が最小となるカメラ配置  $P_c$  を非線形最小化処理により推定する. 以下に非線形最小化の処理手順を示す.

Step1. 自由視点の移動可能範囲  $V$  を取り囲むように初期カメラ配置  $P_c^0$  を設定する.

Step2.  $c$  番カメラについて  $E_{parallax}$  が小さくなる位置に  $P_c$  を微小距離だけ移動する. ただし, カメラを配置不可能な領域や, カメラ群からなる三角形が  $V$  を包含しないような位置にはカメラは移動させないよう拘束を加える (図 59 参照).

Step3. カメラ番号を順に変更する.

Step4.  $E_{parallax}$  の変化がなくなるまで Step2,3 を繰り返す.

以上の最適なカメラ配置の推定処理を室内を模した 3DCG 環境に適用したカメラ配置の推定例を示す. 撮影環境は幅 6m, 奥行き 4m, 高さ 3m の室内を模した環境  $G$  (図 60 中緑線) であり, その環境内の 1m×1m の範囲 (図 60 中青い領域) を自由視点の移動可能範囲  $V$  とした. また, 図 60 中の青い点は,  $E_{parallax}$  の評価を行う自由視点位置  $v$  である.

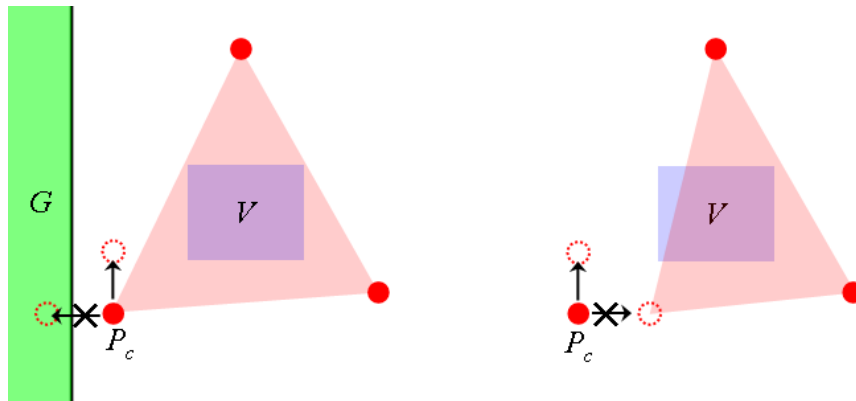


図 59 カメラ位置  $P_c$  の拘束 (左：カメラを配置不可能な領域への移動，右： $V$  を包含しないカメラ配置への更新)

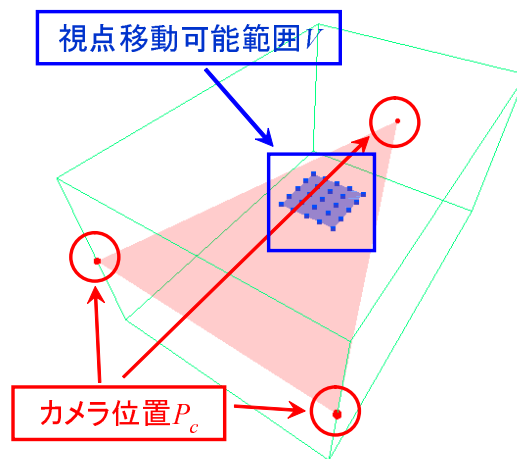


図 60 室内を模した環境  $G$ ，自由視点位置の移動可能範囲  $V$  とカメラ配置  $P_c$

図 61, 62 に  $E_{parallax}$  の最小化処理によるカメラ配置  $P_c$  の変化の様子と，最小化処理の繰り返し回数に対する  $E_{parallax}$  の変化を示す． $E_{parallax}$  の最小化処理が進むにつれ，カメラ間の距離が短くなっていることがわかる．また，この環境と初期カメラ配置では 74 回の繰り返し処理で  $E_{parallax}$  が最小となった．しかし，非線形最小化処理であるため推定されたカメラ配置は初期カメラ配置に依存する．本節で提案したカメラ配置の評価指標は，入力画像と自由視点画像間の視差のみを考慮した指標であり，撮影環境のテクスチャ等は考慮していない．現実環境に

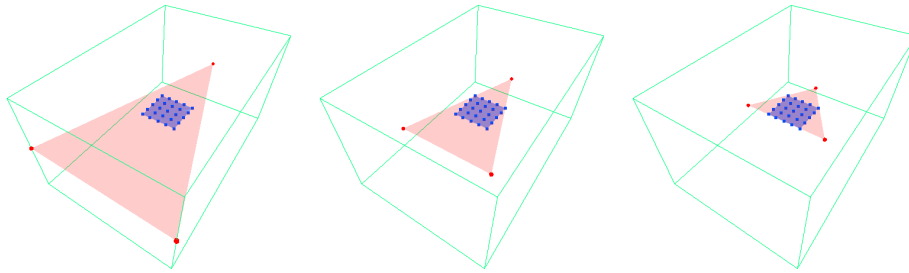


図 61 非線形最小化処理によるカメラ配置  $P_c$  の変化 (左：初期状態，中央：繰り返し回数 37 回，右：繰り返し回数 74 回)

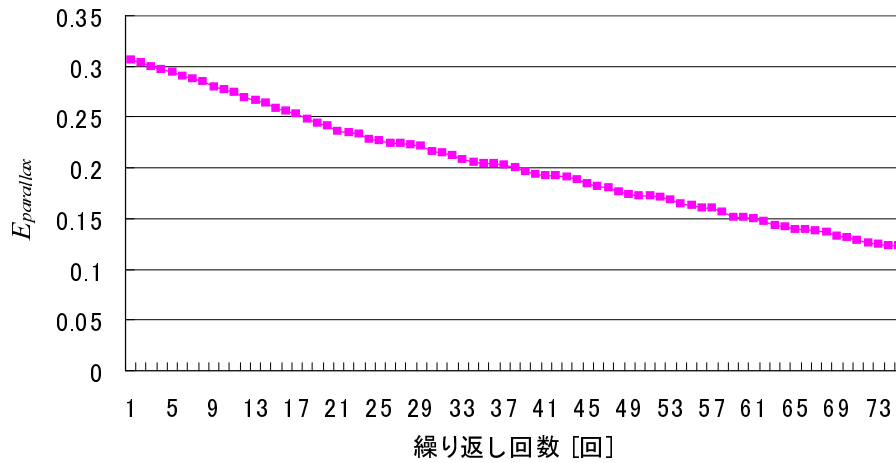


図 62 非線形最小化処理による  $E_{parallelax}$  の変化

において，より利用者の要求に応じたカメラ配置を行うためには，本節で提案した評価指標により推定したカメラ配置からさらに，前節での考察を基にしたカメラ位置の調整を行う必要がある．

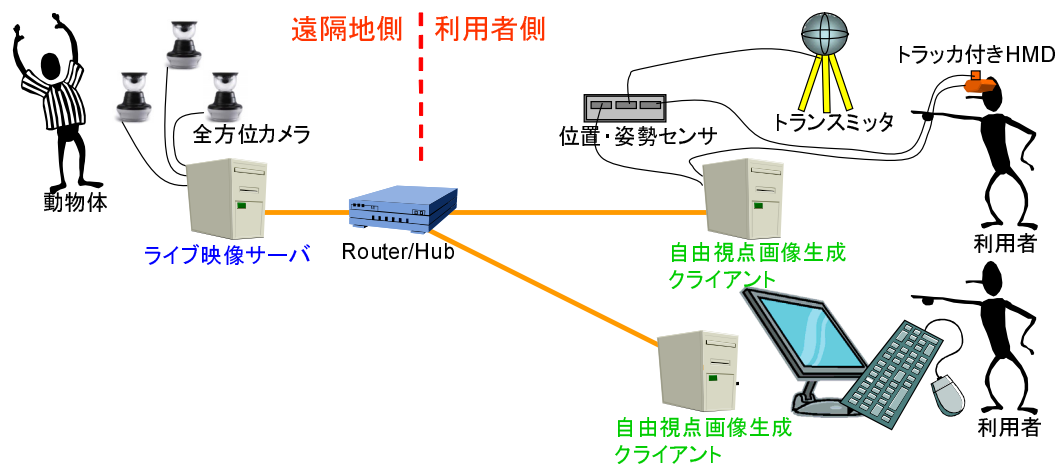


図 63 ハイスケーラブル自由視点テレプレゼンスシステムの構成

## 3.6 ハイスケーラブル自由視点テレプレゼンスシステム

### 3.6.1 システムの設計方針と概要

3.4 節で述べた多視点全方位映像からの自由視点画像生成手法を用い、複数の利用者に自由視点画像を実時間提示可能なハイスケーラブル自由視点テレプレゼンスシステムを提案する。提案システムは、遠隔地で撮影された複数の全方位ライブ映像をネットワーク上でマルチキャスト配信し、利用者はそのライブ映像から生成された自由視点映像を遠隔にいるかのように歩行・見回しをしつつ、高臨場感で観賞可能である。

図 63 に、提案システムの構成を示す。本システムは遠隔で撮影された全方位映像群を配信するライブ映像サーバと配信された全方位映像群から実時間で自由視点画像を生成し、利用者に提示する自由視点画像生成クライアントからなる。ライブ映像サーバは、接続された全方位カメラからの映像をマルチキャストプロトコルを用いてネットワークに配信する。これにより、利用者が増加してもネットワークの負荷が増加しないハイスケーラビリティが実現される。システムの利用者は、HMD と位置姿勢センサからなる視線追従型の映像観賞やマウスやキーボードによるインタラクティブな映像観賞が可能である。

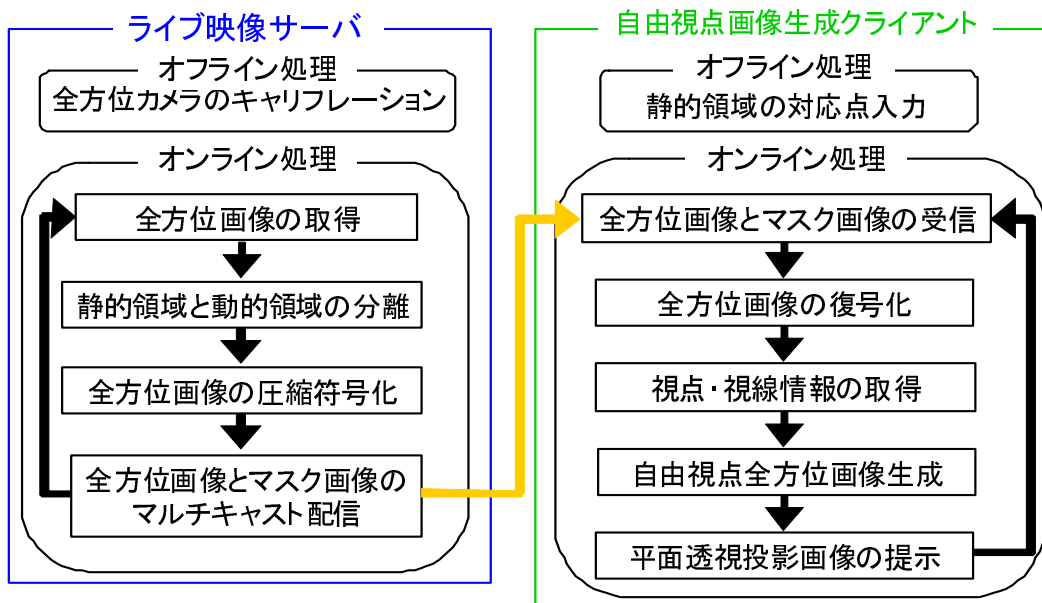


図 64 プロトタイプシステムの処理の流れ

### 3.6.2 サーバ・クライアントの処理内容

本システムでは、ライブ映像サーバと自由視点画像生成クライアントの両方で、事前に計算可能なデータを生成しておくオフライン処理を行うことで、オンライン処理を効率的に行う。さらに、ライブ映像サーバにおいて利用者の視点位置に依存しない背景差分処理を行うことで、自由視点画像生成クライアントの負荷を軽減させる。図 64 に、本システムの処理の流れを示す。また、以下にオフライン・オンライン処理について図 64 に従って説明する。

#### (1) オフライン処理

まず、予め計算可能な以下のオフライン処理をしておく。

[ライブ映像サーバ]

- ・ 環境中の多地点に配置された各全方位カメラの内部・外部パラメータのキャリブレーションを 3.3 節で述べた方法により行う。
- ・ 全方位画像中の 4 隅付近や画像上に投影されたカメラ部は処理する必要のない

無駄な領域なので、これらの領域が処理対象にされないようにマスク画像（2値画像）を生成する。

- ・ 3.4.1 節で述べた背景差分処理により背景画像を生成する。

[自由視点画像生成クライアント]

- ・ 3.4.3 節で述べた静的領域の自由視点画像生成に必要な対応点を予め手動で与える。
- ・ 3.4.4 節で述べた動的領域の自由視点画像生成の際に、自由視点画像中のある画素と自由視点位置を結ぶ直線のベクトル（空間直線ベクトル）が必要になるので、このベクトルは予め自由視点画像上の画素と対応づく状態で計算・保持する。
- ・ ライブ映像サーバと同様に処理する必要のない無駄な領域を計算しないためのマスク画像を生成する。

## (2) オンライン処理

オンライン処理では、時々刻々と取得する全方位動画画像から実時間での自由視点画像生成処理を行う。

[ライブ映像サーバ]

1. 多視点全方位画像を取得し、背景差分処理により動的領域のマスク画像（2値画像）を生成する。
2. 多視点全方位画像を JPEG に圧縮符号化する。
3. 圧縮符号化した多視点全方位画像と動的領域のマスク画像に画像撮影時時刻をタイムコードとして付加したものをパケット化し、マルチキャスト配信する。

[自由視点画像生成クライアント]



1. 圧縮符号化された全方位画像と動的領域のマスク画像を受信し、それらに付加されたタイムコードが多視点画像の全てで一致した場合以下の処理を行う。ただし、図 65 に示すようにクライアントの受信バッファに異なる時刻のタイムコードの画像が入力された場合、古いタイムコードの画像は破棄して新たに多視点画像の受信を待つ。
2. 圧縮符号化された全方位画像を復号化する。
3. 磁気センサやマウス・キーボードから利用者の視線方向・視点位置の情報を取得する。
4. 利用者の視線・視点情報を基に 3.4 節で述べた自由視点画像生成処理を行う。実装上の工夫として、静的領域の自由視点画像生成 (3.4.3 節) の際、三角形パッチ内部の画素に対する画素値のブレンディング処理については、OpenGL を用いて GPU による高速な計算を行う。また、OpenGL による計算で生成された自由視点全方位画像は GPU のフレームバッファ上に存在するので、後の合成処理を高速に行うために PC のメインメモリに戻さず GPU のフレームバッファ上に保持しておく。

動的領域の自由視点画像生成については、3.4.4 節で述べた動物体の存在領域推定に基づく計算量削減に加え、利用者の視線方向の画素のみを Visual Hull の計算対象とする。動的領域の自由視点画像は PC のメインメモリ上に保持されている。後の合成処理を高速に行うため、自由視点画像上で動的領域が存在しない画素についてはアルファ値（不透過度）として 0 を設定する。

静的領域と動的領域の合成処理についても GPU の機能を利用した高速な処理を行う。具体的には、フレームバッファ上に保持されている静的領域の自由視点画像上に、アルファ値（不透過度）が設定された動的領域の自由視点画像をテクスチャとした長方形の平面をレンダリングし、重ね合わせる。

5. 自由視点全方位画像を生成後、生成された全方位画像を平面透視投影変換する。この際、フレームバッファ上の自由視点全方位画像をそのままテクスチャ画

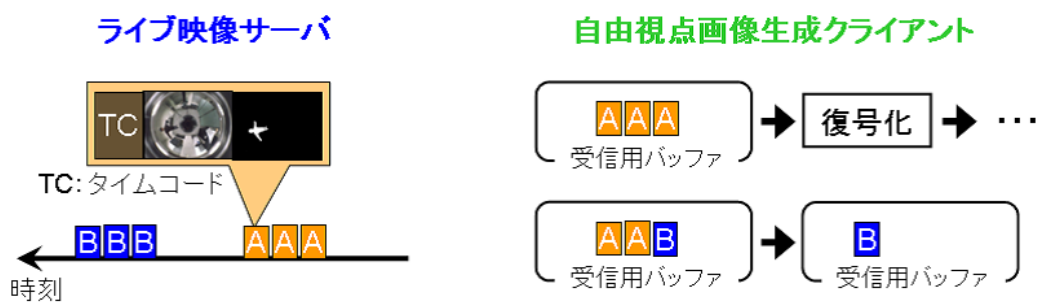


図 65 ライブ映像サーバでのタイムコード付加と自由視点画像生成クライアントでの受信処理

像として利用し，GPU の機能と式 (2) を利用した高速な変換処理 [OYTY98] を行う．

6 . 生成された平面透視投影画像を利用者が装着した HMD やモニタに表示する .

以上の処理を繰り返し行い，1.2 節で述べた A ~ D の機能を満たしたテレプレゼンスを実現する .

### 3.7 複数の利用者による実時間自由視点テレプレゼンス実験

本節では、前節で提案したハイスケーラブル自由視点テレプレゼンスシステムの機能実証実験及び、複数の被験者による主観評価実験について述べる。以下ではまず、提案システムの詳細な機器構成を説明する。次に、提案システムが要求機能 A～D を満たすことを実証するための機能実証実験について述べる。そして、提案システムの有効性を被験者の主観評価により確認する。

#### 3.7.1 システムの構成と実験環境

図 63 に示したハイスケーラブル自由視点テレプレゼンスシステムを実装したプロトタイプシステムについて述べる。本システムは、遠隔地に配置した全方位カメラの画像をマルチキャスト配信する PC (ライブ映像サーバ) と、利用者側で映像を受信し自由視点画像を生成する PC (自由視点画像生成クライアント) からなるサーバ・クライアント型システムである。以下に遠隔地側と利用者側の機器構成を詳述する。

##### [ライブ映像サーバの構成]

IEEE1394 で接続可能な全方位カメラ (未陰産業製 SOIOS 55-cam) 3 台を設置し、それらをすべてライブ映像サーバに接続する。ライブ映像サーバと自由視点画像生成クライアントは、ネットワーク (1Gbps) で接続されている。ライブ映像サーバは、3 台の全方位カメラからの全方位画像と背景差分処理によって得られる動的領域のマスク画像を自由視点画像生成クライアントにマルチキャスト配信する。

##### [自由視点画像生成クライアントの構成]

自由視点画像生成クライアントは、3 視点の全方位画像とそれらに対応する 3 つのマスク画像 (2 値画像) を受信し、自由視点画像を生成する。その際、自由視点画像生成クライアントに接続された電磁気センサ (Polhemus 社製 3SPACE



図 66 電磁気センサ（左）と電磁気レーザ付き HMD（右）

FASTRAK) やマウス・キーボードからの利用者の視線方向・視点位置情報を基に自由視点画像を生成する。利用者の視点位置を比較的広い範囲で計測するために、電磁気トランスミッタにはロングレンジャ(図 66 左)を使用する。システムの利用者は、図 66 右に示す電磁気レーザ付き HMD (オリンパス光学製 FMD-700) を装着し、遠隔地にいるかのように視線方向・視点位置を変え、提示画像を觀賞する。表 4 に使用した全方位カメラとライブ映像サーバ・自由視点画像生成クライアントの詳細を示す。

### 3.7.2 システムの機能実証

プロトタイプシステムにより、ハイスケーラブル自由視点テレプレゼンスの機能実証実験を行った。本実験では、提案システムが実時間ネットワークテレプレゼンスシステムに要求される機能 A~D のすべてを満たすことができることを確認する。そのために、動的環境下で撮影される全方位ライブ映像を用いて複数の利用者が同時に自由視点テレプレゼンスを行い、ネットワークの負荷が増加しないことを確認する。

表 4 プロトタイプシステムの機器構成

全方位カメラ	Resolution: 640x480[画素] Max. frame rate: 15[fps] Field of view: Horizontal:360[度] Vertical:62[度]
ライブ映像サーバ	CPU: Intel Pentium4 3.2GHz
自由視点画像生成クライアント 1	CPU: Intel PentiumD 940 GPU: nVidia GeForce7300GS
自由視点画像生成クライアント 2	CPU: Intel Pentium4 3.2GHz GPU: nVidia GeForce6600GT

システムの利用者は二人とし、一人はHMDと位置・姿勢センサの組み合わせた視線追従型映像提示により遠隔地に居るかのように環境を歩行し、観賞を行う。この利用者を利用者1とする。もう一人は液晶モニタにより提示される映像を観賞し、視点位置の変更をキーボードの矢印キーで、視線方向の変更をマウスのドラッグ操作で行うことができる。この利用者を利用者2とする。本実験の環境は図67左に示す室内環境であり、3つの全方位カメラを配置している。撮影環境中の動物体として、二人の人物に動き回ってもらった。3.3節で述べた全方位カメラの内部・外部パラメータのキャリブレーション手法により推定された各全方位カメラの位置・姿勢を図67右に示す。図中の黒い点はレーザレンジファインダで計測したマーカであり、錐体が推定されたカメラの位置・姿勢を表している。そ

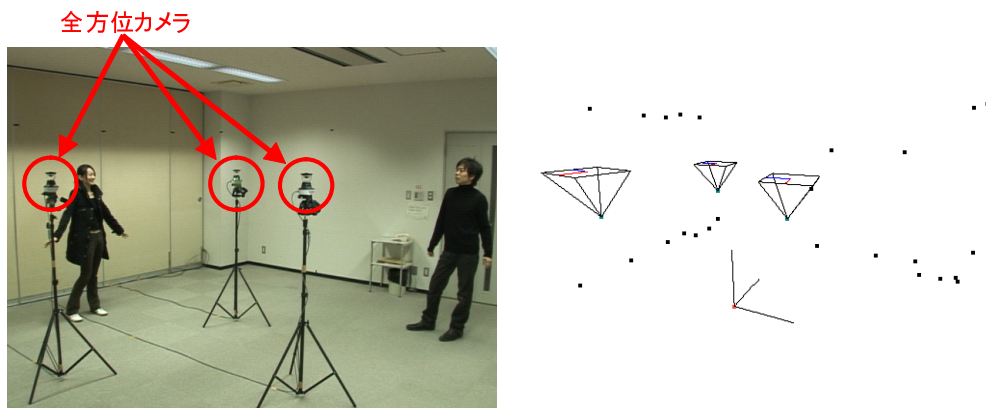


図 67 実験環境（左）と推定されたカメラ位置（右）

それぞれのカメラが概ね正しく推定されていることが分かる．各カメラ間の距離は 2m 程度とし，静的領域の対応点は予め 70 点手動で与えた．またライブ映像サーバは，計算機の性能の制限から 10fps でデータを送信するよう設定した．

図 68,69 に遠隔地と利用者 1 の様子，そして自由視点画像生成クライアント 1 が生成した自由視点全方位画像と各利用者の視線情報を基に平面透視投影変換を行った視界画像を示す．図 68 は動物体を含まず，図 69 は動物体が環境に存在する時の提示画像を示している．利用者 1,2 共に遠隔の動的環境をインタラクティブに観賞することができることを確認した．このとき，各利用者に提示される視界画像の計算には 1 フレームにつき約 120 ~ 200ms 必要であった．その内訳は，マルチキャスト配信された全方位画像と動的領域のマスク画像の取得及び復号化処理に約 80ms，静的領域の自由視点画像生成に約 20ms，動的領域の自由視点画像生成に約 100ms となっている（表 5 参照）．ただし，動的領域の自由視点画像生成は，自由視点画像上に投影される動物体の面積によって大きく処理時間が変動する．図 70 に実験時，ライブ映像サーバがネットワークに送信していたデータの帯域を示す．ライブ映像サーバは，図中の時刻 0 に全方位画像とマスク画像の配信を開始した．その 6 秒後利用者 1 が，24 秒後利用者 2 がテレプレゼンスを開始した．複数の利用者がシステムを利用しているにもかかわらずサーバが送信するデータ量は一定であり，利用者数の増加に対するスケーラビリティを有していることが確認できた．これらの結果から，要求機能 A ~ D のすべての機能が満

表 5 自由視点画像生成クライアントでの処理時間の詳細

全方位画像とマスク画像の受信	約 50ms
全方位画像の復号化	約 30ms
視点・視線情報の取得	約 8ms
静的領域の自由視点画像生成	約 20ms
動的領域の自由視点画像生成	約 100ms
平面透視投影変換	約 1ms

たされていることを確認した。

本実験では、観賞スポットに3台のカメラのみが配置されており、それらのカメラを必ず使用するためシステムの利用者が二人以上になるとネットワーク負荷に関してユニキャストプロトコルよりマルチキャストプロトコルが優位と言える。しかし、観賞スポットに4台以上のカメラが配置されていた場合、そのうち3台のカメラしか使用されない場合も考えられる。このような場合、4台以上のカメラが常に画像をマルチキャストで送信し続けるのはネットワーク帯域の無駄であり、使用するカメラの画像のみを送信するユニキャストの方が優位になる。マルチキャストプロトコルにより画像を配信する場合、使用されていないカメラからは画像を配信するのを停止させる機能を追加することで無駄を解消できると考えられる。

提案システムでは利用者の視点位置・視線方向に追従した視界画像の提示を実現したが、そのフレームレートは約5~7fpsであり滑らかに映像が提示されているとは言い難い。視点・視線に追従した映像を滑らかに提示するには、最低でも10fpsで自由視点画像を生成する必要がある。そのためには、全体の処理時間を約半分にしなければならない。提案した自由視点画像生成手法のうち最も処理時間を要するのは動的領域に対するImage-based Visual Hullである。この処理は、処理対象画素数に比例した計算コストを要する。プロトタイプシステムでは、デュアルコアプロセッサによる並列処理を行っているが、近年普及を始めているクアッドコアプロセッサを利用することでこの処理時間は約半分にすることが可能と考

えられる。また JPEG の復号化処理は現在シングルスレッドで行っているため、この処理についても並列で行うことで、高速化が可能である。静的領域の自由視点画像生成手法に要する約 20ms は、主に多視点画像を GPU のビデオメモリに転送する際にかかる時間である。PC のメインメモリとビデオメモリ間のバスの帯域は現在広帯域化が進んでいるため、近い将来 10ms 以内での転送が可能になると考えられる。以上のことから、計算機の性能向上により自由視点画像生成を 10fps 以上で行うことは近い将来達成可能であると考えられる。

サーバ・クライアント間のネットワーク環境も自由視点画像生成のフレームレートに影響を及ぼす。ネットワークの帯域がシステムに及ぼす影響には、全方位画像とマスク画像の送受信に要する時間と全方位画像の圧縮率の二点が考えられる。全方位画像とマスク画像の送受信に要する時間とネットワーク帯域の関係は次式で表される。

$$\text{送受信時間} = \frac{\text{送受信データ量}}{\text{ネットワーク帯域}} \quad (13)$$

ネットワーク帯域が広いほど送受信時間は短くなり、帯域が狭いほど送受信時間は長くなる。プロトタイプシステムにおいて、サーバ・クライアント間の実効データ転送レートは約 350Mbps であり、自由視点画像 1 フレームを生成するために必要な多視点画像とマスク画像（約 200KB）を伝送するのに要する時間は 5ms 未満であった。ネットワーク帯域が本実験環境の半分の 175Mbps に制限されていた場合、送受信時間はそれぞれ 10ms 以上必要となり、フレームレートが低下を招くことになる。ネットワーク帯域が狭い場合にフレームレートの低下を防止するためには、送受信データ量を少なくする必要がある。提案システムでは、ライブ映像サーバは全方位画像とマスク画像を送信するが、マスク画像については JPEG 等の画像圧縮符号化を行うと動物体の検出結果に誤差が生じるため画像圧縮符号化は行わない。全方位画像に対しては JPEG による画像圧縮符号化を行っているため、この画像圧縮符号化の圧縮率を高めることで送受信データ量を減少させることが可能である。ただし、圧縮率を上げることで画質が劣化するため、システムの利用者は送受信時間とデータ量のバランスを適切に調整する必要がある。

また、データ伝送のジッタ（時間遅延のゆらぎ）や時間遅延、データのロスともフレームレートに影響を及ぼす要因である。プロトタイプシステムでは、3.6.2 節



で述べたようにタイムコードが一致する多視点画像とマスク画像を受信するまでは以降の自由視点画像生成処理を行わない。そして、同一のタイムコードが付加された多視点画像とマスク画像が全て到達する前に、別の時刻のタイムコードが付加された画像が到達するとそれまでに受信した画像群は破棄し、新たに同一のタイムコードが付加された画像群が到達するまで受信待ちを行う。そのため、ネットワーク上で多視点画像のうち一部のデータがロストした場合、その画像群が撮影された時刻のタイムコードを持つデータはすべて破棄され、次の時刻の多視点画像の到達を待つ必要がある。プロトタイプシステムでは、サーバは 10fps で多視点画像とマスク画像を送信しており、データをロストした場合には、約 100ms 後に到達する多視点画像とマスク画像を待つ必要がある。ロストが発生していない場合でも、ジッタによってサーバでのデータの送信順序とクライアントでのデータの受信順序が入れ替わった場合には、データの一部がロストしたと判断し、次の時刻の画像群の受信を待つ。また、本実験では一台のライブ映像サーバが多視点画像を配信しているため、データ伝送の時間遅延は各画像においてジッタの影響はあるもののほぼ同じである。しかし、図 71 のように自由視点画像の生成に必要な多視点画像が異なるサーバから配置される場合、各サーバが配信した画像がクライアントに到達するまでの時間遅延はネットワークの経路によって差が生じる。この差が大きい場合、ジッタによる影響と同様に受信バッファに異なる時刻のタイムコードを持つ画像が入力され、データを破棄するため自由視点画像生成が行われない。これらジッタや時間遅延の影響を受けないためには、本実験環境ではジッタは約 97ms 未満（画像の送信周期 - カメラ 1 台当たりの送信データ量 / ネットワーク帯域  $\times$  (サーバに接続されたカメラ数 - 1)）でなければならない。複数のサーバから配信された画像から自由視点画像を生成する場合は、時間遅延は約 100ms（画像の送信周期）未満でなければならない。また、自由視点画像生成クライアントの受信バッファを大きくし、数フレーム分の多視点画像を蓄積可能にすればこれらの影響を軽減できる。その場合、画像の取得から提示までの時間遅延は大きくなるが、提案システムでは時間遅延の影響を受けることなくインタラクティブな観賞が可能である。

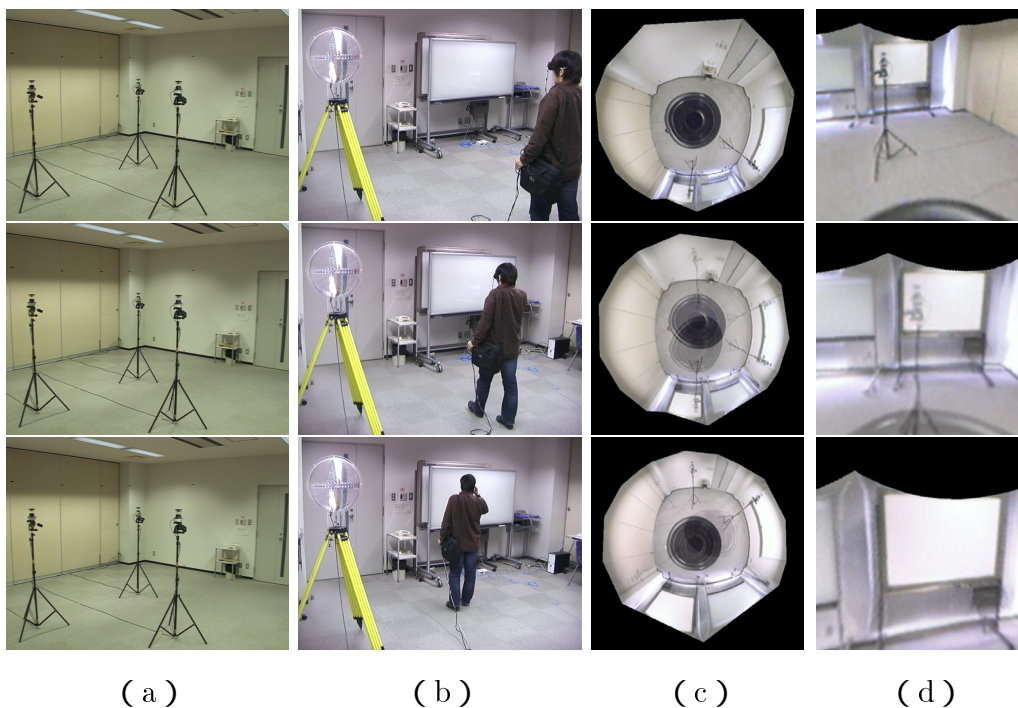


図 68 自由視点テレプレゼンスにおける遠隔地 (a) と利用者 1 の様子 (b) 及び利用者 1 の位置・姿勢情報を基に生成された自由視点全方位画像 (c) と平面透視投影画像 (d)

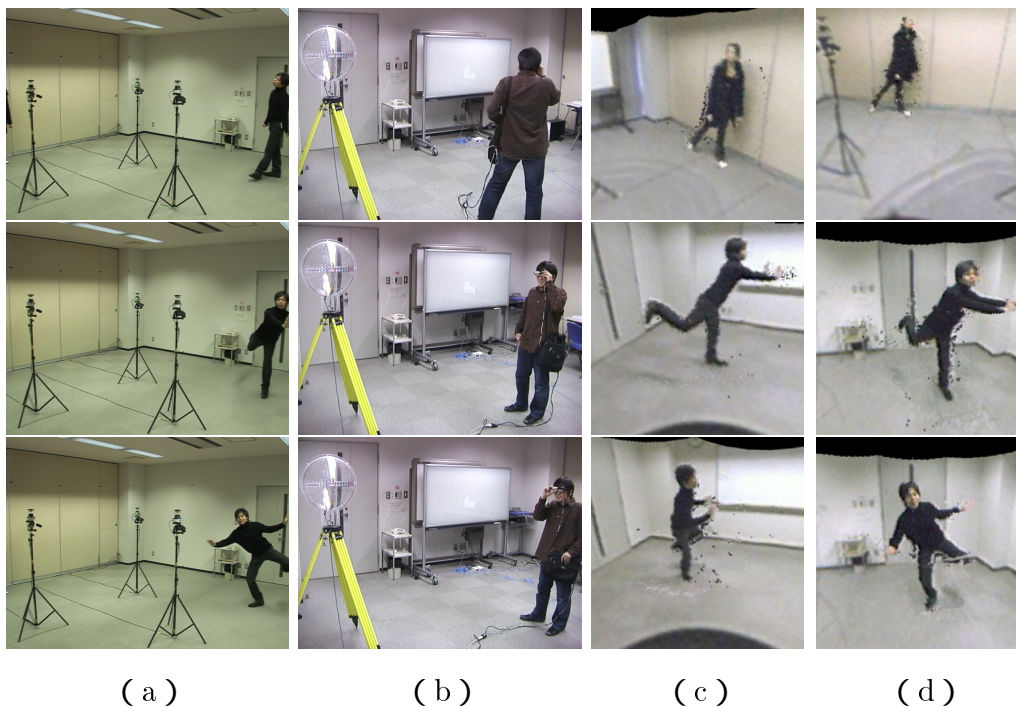


図 69 自由視点テレプレゼンスにおける遠隔地 (a) と利用者 1 の様子 (b) 及び利用者 1,2 の視界画像 (c, d)

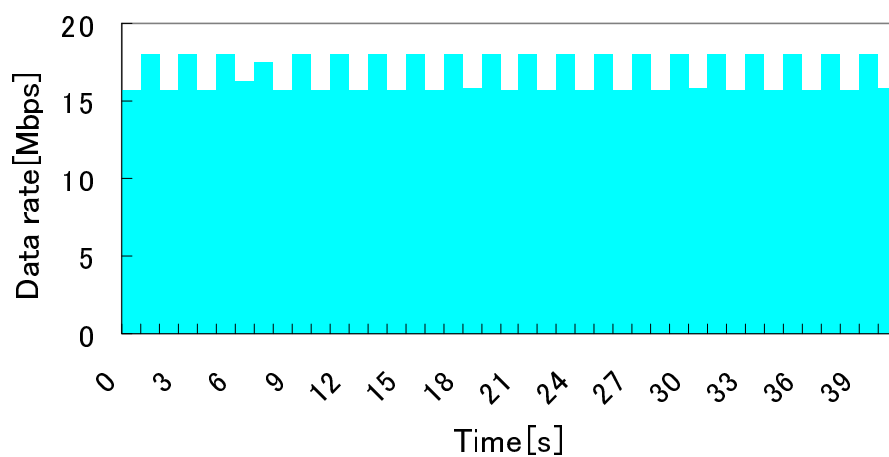


図 70 ライブ映像サーバの動作開始後のネットワーク負荷

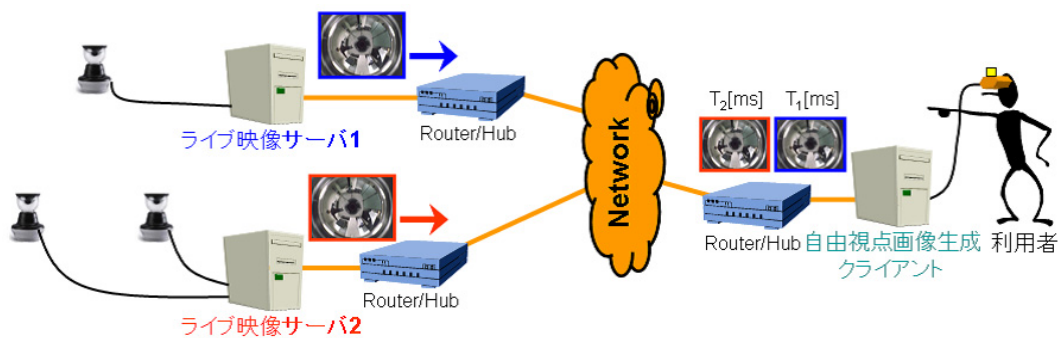


図 71 複数のライブ映像サーバによる多視点画像配信で発生する時間遅延

### 3.7.3 主観評価による有効性の検証

本節では、提案したハイスケーラブル自由視点テレプレゼンスシステムの有効性の確認のために実施した複数の被験者による主観評価実験について述べる。本研究で提案したテレプレゼンスシステムは、従来のシステムに対し高いスケーラビリティと視点位置・視線方向を自由に変更可能なインタラクティブ性を実現している。しかし、3.5.2 節に示したように利用者に提示される自由視点画像は実際にその場で得られる画像に比べ画質が低下することになる。この画質の低下による影響が利用者の動作に応じた映像提示による臨場感向上の妨げになることが考えられる。そこで、提示映像の画質を視点追従性より重視するシステム A と、本研究で提案したシステムそのものである提示映像の視点・視線追従性を画質より重視するシステム B による比較により、画質と視線・視点追従性、そしてそれらを総合して感じられる臨場感において、どちらのシステムが優れるかを複数の被験者により評価してもらった。以下にその評価方法と、その実験の評価結果について述べる。

#### [評価方法]

まず、評価に用いるシステム A と B について述べる。システム A は、提示映像の画質を重視し自由視点画像生成を行わず、入力全方位画像に平面透視投影変換を行った視界画像を利用者に提示する。その際、利用者の視点位置と最も近い実カメラの全方位画像を利用する。システム B は、本研究で提案した自由視点テ

レプレゼンシステムそのものである。両システムの構成は、利用者側に関しては先に述べたプロトタイプシステムと同様であるが（図 63 参照）、実時間での映像の取得は行わず予め撮影し、蓄積した多視点全方位映像を用いて実験を行った。これは、各被験者に提示される仮想環境を同一にすることで、実験条件を合わせる意味がある。

被験者はシステム A と B を体験し、「画質」「視点・視線追従性」「臨場感」のそれぞれの評価項目に関して「A より B がよい」「B より A がよい」「どちらとも言えない」の三択で評価を行う。「画質」は綺麗・自然と感ずるほど、「視線・視点追従性」は動作に応じて提示映像が変化しているほど、「臨場感」は映像に写っている場所にいる感覚が強いほど良い評価を与えるよう各システムの体験前に各被験者に説明した。また、提示映像や被験者の視点位置に関していくつかのパターンを用意し、それぞれのパターン毎にシステム A, B を体験し、各評価項目を評価した。以下にそのパターンを示す。

1. 視点位置を広い範囲で動かした場合（静的環境）

提示される映像中に動物体は含まれず、被験者は 3 つの全方位カメラから成る三角形内をなるべく大きく視点が変化するように動く。

2. 視点位置とカメラとの位置が近い場合（静的環境）

提示される映像中に動物体は含まれず、被験者は実カメラの位置付近で観賞を行う。

3. 視点位置とカメラとの位置が遠い場合（静的環境）

提示される映像中に動物体は含まれず、被験者はどの実カメラからも離れた位置で観賞を行う。

4. 視界画像の更新レートを統一した場合（静的環境）

提示される映像中に動物体は含まれず、被験者はシステム A と B で視界画像の更新レートを統一した映像の観賞を行う。視点位置はパターン 1 同様なるべく大きく視点位置が変化するように動かす。

5. 視点位置を広い範囲で動かした場合（動的環境）

提示される映像中に動物体が存在し，被験者は3つの全方位カメラから成る三角形内をなるべく大きく視点が変化するように動く．

6. 視点位置とカメラとの位置が近い場合（動的環境）

提示される映像中に動物体が存在し，被験者は実カメラの位置付近で観賞を行う．

7. 視点位置とカメラとの位置が遠い場合（動的環境）

提示される映像中に動物体が存在し，被験者はどの実カメラからも離れた位置で観賞を行う．

8. 視界画像の更新レートを統一した場合（動的環境）

提示される映像中に動物体が存在し，被験者はシステム A と B で視界画像の更新レートを統一した映像の観賞を行う．視点位置はパターン 1 同様になるべく大きく視点位置が変化するように動かす．

被験者は各パターンについて各システムを約 1 分間体験する．

本実験で被験者に提示した映像は，本学ミレニアムホール前の屋外環境である（図 72 参照）．被験者として，本研究で提案するシステムに関する詳しい知識を持たず，本学のミレニアムホールに行ったことのある 12 名に参加してもらった．本学ミレニアムホールに行った経験があることにより，各システムを体験した際の臨場感に関する評価が正しく行われることが期待できる．

[評価結果]

表 6 に 12 名の被験者による評価結果を示す．これらのパターン・評価項目のうち，有意水準 5% のもとで「A より B がよい」「B より A がよい」「どちらとも言えない」の三択間に統計的有意な差が確認された項目は赤色で示している．画質に関しては，全てのパターンにおいて提案システムである B よりもシステム A が良いという結果となった．このことから被験者には，はっきりとシステム B が生成した自由視点画像が入力画像よりも劣ると判断できることが分かる．特に，環境に動物体が存在しているパターン 5~8 では，B よりも A が高評価であることが



図 72 実験時に被験者に提示された映像が取得された環境（本学ミレニアムホール前）

顕著に現れている．これは動物体の自由視点画像生成の際にアーティファクトが発生してしまうことが原因と考えられる．視点・視線追従性に関しては，環境に動物体が存在し視点位置を大きく動かしたパターンにおいて，提案システムであるシステム B が良いという結果を得ている．これは動物体が比較の実カメラに近い位置に存在しており，システム A ではカメラの切り替わりによって動物体の居る方向が大きく変化してしまい，動物体が消えたかのように感じてしまうことが原因として考えられる．これに対し，システム B では視点位置の変化に対して連続な映像提示が行えるため，そのような違和感を感じることはなかったことが高評価に繋がったと考えられる．臨場感に関しては，動的環境であまり視点を動かさなかった場合にはシステム A が良いという結果になっている．これは，視点を動かさないため，システム B の自由視点画像生成の効果がほとんど影響しなかったため画質の良いシステム A が総合的には良いと判断されたと考えられる．しかし，パターン 8 においては，システム B の方が臨場感が高いという結果になっている．システム A の方が自由視点画像生成を行わないため映像提示の更新レートが高いが，システム A にシステム B と同様な計算負荷を与えることで更新レートを統一すると，統一前よりシステム A の視線方向に関する追従性が低下するため，総合的にみてシステム B の方が高評価になったと考えられる．今後，コンピュータの高性能化により，システム B の提示映像の更新レートも高くなることが予想されるため，システム A よりも高臨場感となる可能性がある．

表 6 12名の被験者による評価結果（単位：人）

		A > B	A = B	A < B	
静的環境	1. 視点位置を広い範囲で動かした場合	画質	9	3	0
		視線・視点追従性	3	3	6
		臨場感	6	1	5
	2. 視点位置とカメラとの位置が近い場合	画質	6	6	0
		視線・視点追従性	8	2	2
		臨場感	6	5	1
	3. 視点位置とカメラとの位置が遠い場合	画質	10	1	1
		視線・視点追従性	3	4	5
		臨場感	6	1	5
	4. 視界画像の更新レートを統一した場合	画質	10	2	0
		視線・視点追従性	4	1	7
		臨場感	7	1	4
動的環境	5. 視点位置を広い範囲で動かした場合	画質	11	0	1
		視線・視点追従性	1	1	10
		臨場感	5	4	3
	6. 視点位置とカメラとの位置が近い場合	画質	12	0	0
		視線・視点追従性	3	9	0
		臨場感	12	0	0
	7. 視点位置とカメラとの位置が遠い場合	画質	11	0	1
		視線・視点追従性	4	5	3
		臨場感	9	2	1
	8. 視界画像の更新レートを統一した場合	画質	11	0	1
		視線・視点追従性	1	1	10
		臨場感	2	1	9



### 3.8 結言

本章では、前章で提案したハイスケーラブルな実時間ネットワークテレプレゼンスシステムを基に、多視点全方位映像からの実時間自由視点画像生成処理を加えたハイスケーラブル自由視点テレプレゼンスシステムを提案した。そして、提案システムのプロトタイプを用いた機能実証実験により、要求機能 A~D を全て満たすことが可能であることを示した。また、複数の被験者による主観評価実験により、提案システムの有効性を示した。

本章で提案したハイスケーラブル自由視点テレプレゼンスシステムは、画像の圧縮符号化に JPEG を利用しており、少ない時間遅延で遠隔の環境を観賞可能である。そのため、本システムを保育施設に適用することで複数の親が同時に現在の子供の様子を見ることができシステムが実現可能である。ただし、そのようなシステムに適用する場合には Web ブラウザ等で簡単に観賞が可能なインタフェースにすることなどの検討が必要になる。また、本システムを監視システムに適用することで、カメラのスイッチングによるカメラ映像の切り替えを行うシステムに比べ、監視環境と監視カメラの位置関係が把握しやすいシステムを実現できると考える。

機能実証実験において利用者に提示された画像は、画質評価実験と同様に、自由視点位置が全方位カメラに近いほど、良好な結果となった。しかし、機能実証実験では動物体がほぼ直立状態であったため、画質評価実験のような実際の物体の形状と Visual Hull の形状の差が表れにくくなっており、自由視点位置が動物体に近い場合にもあまり画質の低下を感じることはなかった。画質評価実験の静的環境は、撮影対象とカメラとの距離が比較的離れていたため各カメラの画像上で同様に映る部分が多く、自由視点画像を生成した際にも大きな歪みは見られなかったが、本実験環境では撮影対象とカメラとの距離が短く、各カメラで同様に撮影できない部分がいくつか存在し、対応点が与えられない部分があった。その影響で、自由視点位置がカメラから離れた場合に、対応点が与えられていない部分に大きな歪みが発生していた。また、相互のカメラが各全方位画像中に映っているため、自由視点位置において半透明なカメラが設置されているように見えることがあった。これは、Morphing の際の画像のブレンディングによる影響であ

るが、別のカメラが映っている領域に対しては、ブレンディングの比率を変更することで、この映りこみは減少させられると考える。

## 4. 結論

### 4.1 本論文のまとめ

本論文では、遠隔の観光地や名所の風景をインタラクティブに観賞するようなアプリケーションへの応用を想定し、複数の利用者が遠隔地の動的環境中を実時間かつ自由な視線方向・視点位置で観賞可能なハイスケーラブル自由視点テレプレゼンスを提案した。提案システムは、実時間ネットワークテレプレゼンスシステムに要求される以下の A～D の機能を全て満たすシステムである。

- A. 全方位の動的環境の仮想化：撮影環境中の多地点に全方位カメラを配置し、同期撮影することにより実現。
- B. 映像の実時間ネットワーク伝送：多地点全方位映像をネットワークを介して伝送することで実現。
- C. 利用者の動作に応じたインタラクティブな映像提示：多視点全方位映像からの実時間自由視点画像生成により実現。
- D. 利用者数に対するスケーラビリティ：利用者の視点位置・視線方向に依存しない全方位映像のマルチキャスト配信により実現。

本研究では、この自由視点テレプレゼンスシステムの実現のために、従来の全方位映像を用いたネットワークテレプレゼンスシステムに関して問題とされていた、要求機能 C の視点位置に応じた映像提示と要求機能 D の利用者数に対するスケーラビリティの改善を行った。従来のシステムは、遠隔地の映像をネットワーク伝送することには対応していたが、ユニキャストによる映像配信を行っていたため、利用者数に比例したネットワーク帯域とサーバの処理コストが必要であった。これに対し、全方位映像は利用者の視線方向には依存しないという特徴を利用して、それらをマルチキャストプロトコルを用いて配信することにより、利用者数が増加してもネットワーク帯域とサーバの処理コストの増加しない実時間ネットワークテレプレゼンスシステムを実現した。これにより要求機能 D を満たした。さらに、利用者のユーザビリティも考慮した Web ブラウザベース全方位映像

ビューアにより，誰でも簡単に全方位映像コンテンツの観賞が可能なシステムとした．

この実時間ネットワークテレプレゼンスシステムの残された課題であった，要求機能 C の利用者の視点位置に応じた映像提示を，多視点全方位映像からの実時間自由視点画像生成により解決した．撮影環境中の多地点に配置した全方位カメラにより同期撮影することで，動的環境の仮想化に対応し，広範囲の動的環境において実時間で自由視点画像を生成するため，撮影環境を静的領域と動的領域に分離し，それぞれの領域に対して適した自由視点画像生成手法を適用した．また，処理コストの高い動的領域に対しては，動的領域の存在領域推定を併用した高速な自由視点画像生成を実現した．この自由視点画像生成技術を実時間ネットワークテレプレゼンスシステムに組み込み，ハイスケーラブル自由視点テレプレゼンスシステムを実現した．そして，本システムを用いた機能実証実験により，要求機能 A～D が全て満たされることを確認した．また，複数の被験者による主観評価実験により，自由視点画像生成による利用者の動作に追従した映像提示が，視点追従性の向上及び高臨場感の実現に有効であることを確認した．

## 4.2 今後の検討課題

今後の検討課題には，自由視点画像の画質とフレームレートの向上，システムをより広域の空間に対応させるためのスケラビリティの実現が挙げられる．定量的画質評価や主観評価実験による結果からも，入力画像に対して生成した自由視点画像の画質は低下しており，これが臨場感を損なう一因と考えられる．また，自由視点画像生成の計算コストにより従来の全方位映像を用いたテレプレゼンスシステムに比べフレームレートが低下していることも，臨場感を損なう原因と考えられる．自由視点画像の画質低下の原因としては，全方位カメラの内部・外部パラメータの推定精度が不足していることが考えられる．本研究での全方位カメラ HyperOmni Vision のキャリブレーションは，双曲面ミラーの回転軸とカメラの光軸が一致していることを仮定し，パラメータを推定しているが，実際の全方位カメラは，ミラーとカメラの設置不良によりそれらは一致していない．そのため推定した各パラメータに誤差を生じ，自由視点画像生成にその誤差の影響が現

れたと考えられる．よって，今後はより高精度なキャリブレーション手法を用いることを検討すべきである．また自由視点画像生成処理についても，生成した自由視点画像にアーティファクトが発生しているかどうかを自動的に判断する処理を加え，それが発生していた場合にはより自然な画像となるよう修正するような機能が必要である．このような処理は非常に高い計算コストを要するが，マルチコア CPU や GPU の利用により近い将来実時間でも実行可能になると考える．またフレームレートについても計算機の高性能化により改善が期待できる．

システムをより広域の空間に対応させるためのスケーラビリティに関して，今後あらゆる場所にネットワーク接続されたカメラが配置されることが想定されるため，それらを利用して遠隔のあらゆる場所を観賞可能なシステムが実現できると考えられる．そのためには，全方位カメラだけでなく通常のカメラや魚眼レンズによる広角なカメラも統一的に利用可能なフレームワークが必要である．あらゆる種類のカメラを利用可能なシステムとすることで，撮影環境に適したカメラを配置することが可能となり，自由視点画像生成に利用可能な映像が増えることにつながり，より広範囲の視点移動が可能なテレプレゼンスシステムが実現可能と考える．

## 謝辞

全過程を通して細やかな御指導を賜りました，視覚情報メディア講座 横矢 直和 教授に心より感謝致します．特に，研究ミーティングや発表論文の添削においては，数々の印象に残る御言葉を頂きました．また，横矢教授は国内外において研究成果を発表する多くの機会を与えて下さりました．

本研究に関して，御助言，御鞭撻頂いたロボティクス講座 小笠原 司 教授に厚く御礼申し上げます．

本研究の遂行にあたり，御助言，御鞭撻頂いたインターネット・アーキテクチャ講座 砂原 秀樹 教授，藤川 和利 准教授に厚く御礼申し上げます．また博士前期課程において，実験環境の整備や研究に関しての御助言を頂いた inet-lab 中村 豊 助手（現，九州工業大学 准教授），佛圓 裕一 君に深く感謝致します．

奈良先端科学技術大学院大学 21 世紀 COE プログラム「ユビキタス統合メディアコンピューティング」拠点リーダー 千原 国宏 教授をはじめ，本研究を進めるにあたり有益な御助言や励ましの言葉を頂いた同プログラムのコアリーダーおよび各研究グループの諸先生方，事務補佐員の皆様，研究員の皆様に深く感謝いたします．

そして，本研究の全過程を通して温かい御指導をして頂いた視覚情報メディア講座 山澤 一誠 准教授に深く感謝申し上げます．物心両面において熱く御支援を頂いた視覚情報メディア講座 神原 誠之 助教に深く御礼申し上げます．研究活動の全過程を通して多くの御助言，御指導賜りました視覚情報メディア講座 佐藤 智和 助教に心より感謝致します．日々，研究室での研究活動を支えてくださった視覚情報メディア講座事務補佐員 守屋 知代，高橋 美央，両女史に心より感謝申し上げます．研究活動において常に温かい御助言を頂いた視覚情報メディア講座の諸氏に感謝いたします．特に，本研究の実験に御助力下さり，また共同研究においては熱心に研究に取り組み，研究の楽しさを再認識させて下さった大野 純佳 氏に感謝いたします．

最後に，学業に専念する機会を与えて下さり，多くの面で支えて下さった両親と兄弟に深く感謝します．

## 参考文献

- [Bak77] H. Baker, "Three-dimensional modeling," *In Proc. of 5th Int. Joint Conf. on Artificial Intelligence*, pp. 649-655, 1977.
- [CB61] C. Comeau and J. Bryan, "Headsight television system provides remote surveillance," *Electronics*, pp. 86-90, 1961.
- [Che95] S. E. Chen, "QuickTime VR - an image-based approach to virtual environment navigation," *In Proc. of SIGGRAPH'95*, pp. 29-38, 1995.
- [COH85] Z. L. Cao, S. J. Oh and E. L. Hall, "Dynamic omnidirectional vision for mobile robots," *In Proc. of SPIE Intelligent Robots and Computer Vision*, Vol. 579, pp. 405-414, 1985.
- [Cor02] Microsoft Corporation, "Windows media encoder 9 series," <http://www.microsoft.com/windows/windowsmedia/9series/encoder/default.aspx>, 2002.
- [CW96] S. E. Chen and L. Williams, "View interpolation for image synthesis," *In Proc. of SIGGRAPH'96*, pp. 279-288, 1996.
- [Ear96] EarthCam, Inc., "Earthcam -webcam network-," <http://www.earthcam.com/>, 1996.
- [FWZ05] A. Fitzgibbon, Y. Wexler and A. Zisserman, "Image-based rendering using image-based priors," *Int. Journal of Computer Vision (IJCV)*, Vol. 2, pp. 141-151, 2005.
- [GGSC96] S. J. Gortler, R. Grzeszczuk, R. Szeliski and M. F. Cohen, "The lumigraph," *In Proc. of SIGGRAPH'96*, pp. 43-54, 1996.
- [Hec94] P. Heckbert, "Graphics games IV," *Academic Press*, pp. 47-59, 1994.

- [IHA03] M. Irani, T. Hassner and P. Anandan, "What does the scene look like from a scene point?," *In Proc. of 7th European Conf. on Computer Vision (ECCV2003)*, pp. 1176-1183, 2003.
- [KKO03] T. Koyama, I. Kitahara and Y. Ohta, "Live mixed-reality 3D video in soccer stadium," *In Proc. of 2nd IEEE/ACM Int. Symposium on Mixed and Augmented Reality (ISMAR03)*, pp. 987-990, 2003.
- [KNR97] T. Kanade, P. J. Narayanan and P. Rander, "Virtualized Reality: Constructing virtual worlds from real scenes," *IEEE Multimedia*, Vol. 4, No. 1, pp. 34-47, 1997.
- [KO03] I. Kitahara and Y. Ohta, "Scalable 3D representation for 3D video display in a large-scale space," *In Proc. of IEEE Virtual Reality 2003 Conf. (VR2003)*, pp. 45-52, 2003.
- [KYI<sup>+</sup>98] K. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura and N. Yokoya, "Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors," *In Proc. of 14th IAPR Int. Conf. on Pattern Recognition (ICPR'98)*, pp. 485-489, 1998.
- [LH96] M. Levoy and P. Hanrahan, "Light field rendering," *In Proc. of SIGGRAPH'96*, pp. 31-42, 1996.
- [Lip80] A. Lippman, "Movie Maps: An application of the optical videodisc to computer graphics," *In Proc. of SIGGRAPH'80*, pp. 32-43, 1980.
- [MBR<sup>+</sup>00] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler and L. McMillan, "Image-based visual hulls," *In Proc. of SIGGRAPH2000*, pp. 369-374, 2000.
- [Min80] M. Minsky, "Telepresence," *OMNI Magazine*, Vol. 2, pp. 45-52, 1980.



- [Moe97] S. Moezzi, "Immersive telepresence," *IEEE MultiMedia*, Vol. 4, No. 1, pp. 17-56, 1997.
- [MT02] T. Matsuyama and T. Takai, "Generation, visualization, and editing of 3D video," *In Proc. of 1st Int. Symposium on 3D Data Processing, Visualization and Transmission*, pp. 234-245, 2002.
- [OK93] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 15, No. 4, pp. 353-363, 1993.
- [OYTY98] Y. Onoe, K. Yamazawa, H. Takemura and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams," *Computer Vision and Image Understanding (CVIU)*, Vol. 71, No. 2, pp. 154-165, 1998.
- [Par00] Paris-Live.com, "Eiffel tower webcams -paris france video-news-," <http://www.paris-live.com/>, 2000.
- [SBK03] H. Saito, S. Baba and T. Kanade, "Appearance-based virtual view generation from multicamera videos captured in the 3-D room," *IEEE Trans. on Multimedia*, Vol. 5, No. 3, pp. 303-316, 2003.
- [SD96] S. M. Seitz and C. R. Dyer, "View morphing," *In Proc. of SIGGRAPH'96*, pp. 21-30, 1996.
- [SD98] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," *In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'98)*, pp. 1067-1073, 1998.
- [STYY00] J. Shimamura, H. Takemura, N. Yokoya and K. Yamazawa, "Construction and presentation of a virtual environment using panoramic stereo images of a real scene and computer graphics models," *In Proc.*

- of *15th Int. Conf. on Pattern Recognition (ICPR2000)*, pp. 463-467, 2000.
- [Sut65] I. Sutherland, "The ultimate display," *In Proc. of IFIPS Congress 1965*, Vol. 2, pp. 506-508, 1965.
- [Tay02] C. J. Taylor, "VideoPlus: A method for capturing the structure and appearance of immersive environment," *IEEE Trans. on Visualization and Computer Graphics*, Vol. 8, No. 2, pp. 171-182, 2002.
- [TTY02] K. Tomite, K. Yamazawa and N. Yokoya, "Arbitrary viewpoint rendering from multiple omnidirectional images," *In Proc. of 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002)*, pp. 987-990, 2002.
- [UCK<sup>+</sup>04] M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Hartley and R. Szeliski, "High-quality image-based interactive exploration of real-world environments," *IEEE Computer Graphics and Applications*, Vol. 24, No. 3, pp. 52-63, 2004.
- [YD01] T. Yamamoto and M. Doi, "PanoVi: Panoramic movie system for real-time network transmission," *In Proc. of IEEE 4th Workshop on Multimedia Signal Processing*, pp. 389-394, 2001.
- [YIS04] S. Yamazaki, K. Ikeuchi and Y. Shingawa, "Determining plausible mapping between images without a priori knowledge," *In Proc. of Asian Conference on Computer Vision 2004 (ACCV2004)*, pp. 408-413, 2004.
- [YNY91] Y. Yagi, Y. Nishizawa and M. Yachida, "Estimating location and avoiding collision against unknown obstacle for the mobile robot using omnidirectional image sensor copic," *In Proc. of Int. Workshop on Intelligent Robots and System*, pp. 909-914, 1991.

- [YSK<sup>+</sup>02] S. Yamazaki, S. Sagawa, H. Kawasaki, K. Ikeuchi and M. Sakauchi, “Microfacet billboarding,” *In Proc. of the 13th Eurographics Workshop on Rendering*, Vol. 13, pp. 169-179, 2002.
- [YY03] K. Yamazawa and N. Yokoya, “Detecting moving objects from omnidirectional dynamic images based on adaptive background,” *In Proc. of 10th IEEE Int. Conf. on Image Processing (ICIP2003)*, Vol. III, pp. 953-956, 2003.
- [ZKU<sup>+</sup>04] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder and R. Szeliski, “High-quality video view interpolation using a layered representation,” *ACM Trans. on Graphics*, Vol. 23, No. 3, pp. 600-608, 2004.
- [浅井 05] 浅井俊弘, 神原誠之, 横矢直和, “全方位距離画像と全方位カメラ画像の統合による屋外環境の三次元モデル化”, *画像電子学会誌*, Vol. 34, No. 5, pp. 529-538, 2005.
- [池田 03] 池田聖, 佐藤智和, 横矢直和, “全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画の生成とテレプレゼンスへの応用”, *日本バーチャルリアリティ学会論文誌*, Vol. 8, No. 4, pp. 443-450, 2003.
- [稲本 04] 稲本奈穂, 斎藤英雄, “視点位置の内挿に基づく3次元サッカー映像の自由視点鑑賞システム”, *映像情報メディア学会誌*, Vol. 58, No. 4, pp. 529-539, 2004.
- [上田 05] 上田恵, 有田大作, 谷口倫一郎, “多視点動画処理による3次元モデル復元に基づく自由視点画像生成のオンライン化-pc クラスタを用いた実現法-”, *情報処理学会論文誌*, Vol. 46, No. 11, pp. 2768-2778, 2005.
- [ウ 01] ウ小軍, 和田俊和, 東海彰吾, 松山隆司, “平面間透視投影を用いた並列視体積交差法”, *情報処理学会論文誌: コンピュータビジョンとイメージメディア*, Vol. 42, No. SIG6(CVIM2), pp. 33-43, 2001.

- [エヌ 02] エヌ・ティ・ティ ネオメイト, “みーんな 元気だよ”, <http://www.ntt-neo.com/news/2002/020726.html>, 2002.
- [遠藤 02] 遠藤隆明, 片山昭宏, 田村秀行, 廣瀬通孝, “写実的な広域仮想空間構築のための画像補完手法”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 185-192, 2002.
- [岡 03] 岡慎也, ナ・バンチャンプリム, 藤井俊彰, 谷本正幸, “自由視点テレビのための光線空間情報圧縮”, 電子情報通信学会 技術研究報告 IE2003-131, 2003.
- [小野 06] 小野定康, “遠隔医療”, 映像情報メディア学会学会誌, Vol. 60, No. 6, pp. 853-856, 2006.
- [川西 06] 川西隆仁, 田辺暁弘, 松井健一, 深津真二, 堀越崇, 森永徹, 伊藤康之, 中岡聡, “Ip 映像配信サービス「4th media」”, 映像情報メディア学会学会誌, Vol. 60, No. 6, pp. 847-852, 2006.
- [高橋 06] 高橋桂太, 苗村健, “視点依存奥行きマップ実時間推定に基づく多眼画像からの自由視点画像合成”, 映像情報メディア学会誌, Vol. 60, No. 10, pp. 1611-1622, 2006.
- [田中 06] 田中啓仁, “Ftth サービス技術の現状と将来展望”, 映像情報メディア学会学会誌, Vol. 60, No. 6, pp. 860-864, 2006.
- [棚橋 02] 棚橋英樹, 佐藤淳, 王彩華, 丹波義典, 山本和彦, “全方位ステレオシステム (sos) のキャリブレーション”, 映像情報メディア学会誌, Vol. 56, No. 4, pp. 603-610, 2002.
- [富手 02] 富手要, 山澤一誠, 横矢直和, “複数の全方位画像を用いた広範囲なウォークスルーの実現”, 画像の認識・理解シンポジウム (MIRU2002) 講演論文集 Vol. II, pp. 353-358, 2002.

- [根岸 04] 根岸善郎, 子安大士, 三浦純, 白井良明, “移動ロボットのための全方位ステレオのキャリブレーション”, 2004 年ロボティクス・メカトロニクス講演会, Vol. 2P2-L1-49, , 2004.
- [ビュー03] ビュー プ ラ ス, “LiveSPHERE,” <http://www.viewplus.co.jp/products/livesphere/index.html>, 2003.
- [北海 99] 北 海 道 新 聞 社, “北 海 道新聞 時計台ライブカメラ”, <http://yoho.hokkaido-np.jp/tokeidai/>, 1999.
- [森田 02] 森田真司, 山澤一誠, 横矢直和, “全方位画像センサによるネットワークを介したテレプレゼンスシステムと遠隔監視システム”, 電子情報通信学会 技術報告 PRMU2002-83, 2002.
- [森田 05] 森田真司, 山澤一誠, 寺沢征彦, 横矢直和, “全方位画像センサを用いたネットワーク対応型遠隔監視システム”, 電子情報通信学会論文誌 (D-II) , Vol. J88-D-II, No. 5, pp. 864-875, 2005.
- [山碕 06] 山碕良志, “我が国 ict 政策の動向と展望”, 映像情報メディア学会学会誌, Vol. 60, No. 6, pp. 840-846, 2006.
- [山口 01] 山口晃一郎, 山澤一誠, 竹村治雄, 横矢直和, “全方位動画像を用いた両眼ステレオ画像の実時間生成によるテレプレゼンス”, 電子情報通信学会論文誌 (D-II), Vol. 84, No. 9, pp. 2048-2057, 2001.
- [山澤 96] 山澤一誠, 八木康史, 谷内田正彦, “移動ロボットのナビゲーションのための全方位視覚センサ HyperOmni Vision の提案”, 電子情報通信学会論文誌 (D-II) , Vol. J79-D-II, No. 5, pp. 698-707, 1996.
- [山澤 02] 山澤一誠, 横矢直和, “全方位カメラと JAVA applet を用いたインターネット定点カメラ”, 情報科学技術フォーラム (FIT) 一般講演論文集, Vol. 3, No. I-7, pp. 13-14, 2002.

## 研究業績

### 学術論文誌

1. 石川 智也, 山澤 一誠, 横矢 直和, "複数の全方位動画像を用いた自由視点テレプレゼンス", 映像情報メディア学会誌, Vol. 60, No. 10, pp. 1599–1610, 2006. (第3章に関連)
2. 山澤 一誠, 石川 智也, 中村 豊, 藤川 和利, 横矢 直和, 砂原 秀樹, "Webブラウザと全方位動画像を用いたテレプレゼンスシステム", 電子情報通信学会論文誌 (D-II), Vol. J88-D-II, No. 8, pp. pp. 1750–1753, 2005. (第2章に関連)

### 国際会議

1. T. Ishikawa, K. Yamazawa, and N. Yokoya, "Novel view telepresence with high-scalability using multi-casted omni-directional videos," *In Proc. of 9th Int. Conf. on Enterprise Information Systems (ICEIS2007)*, Vol. HCI, pp. 148–156, 2007. (第3章に関連)
2. T. Ishikawa, K. Yamazawa, and N. Yokoya, "Novel view telepresence using multiple omni-directional live videos," *In Proc. of ACM SIGGRAPH2006*, Research poster, 2006. (第3章に関連)
3. T. Ishikawa, K. Yamazawa, and N. Yokoya, "Novel view generation from multiple omni-directional videos," *In Proc. of IEEE and ACM Int. Sympo. on Mixed Augmented Reality (ISMAR05)*, pp. 166–169, 2005. (第3章に関連)
4. T. Ishikawa, K. Yamazawa, and N. Yokoya, "Real-time generation of novel views of a dynamic scene using morphing and visual hull," *In Proc. of 12th IEEE Int. Conf. on Image Processing (ICIP2005)*, Vol. I, pp. 1013–1016, 2005. (第3章に関連)

5. K. Fujikawa, K. Yamazawa, T. Ishikawa, Y. Butsuen, T. Sato, Y. Nakamura, H. Sunahara, and N. Yokoya, "A networked telepresence system for cars and its transmission issue," *In Proc. of 11th Annual Euromedia Conference 2005*, pp. 119–123, 2005. (第2章に関連)
6. T. Ishikawa, K. Yamazawa, T. Sato, S. Ikeda, Y. Nakamura, K. Fujikawa, H. Sunahara, and N. Yokoya, "Networked telepresence system using web browsers and omnidirectional video streams," *In Proc. of SPIE Electronic Imaging*, Vol. 5664, pp. 380–387, 2005. (第2章に関連)
7. K. Yamazawa, T. Ishikawa, T. Sato, S. Ikeda, Y. Nakamura, K. Fujikawa, H. Sunahara, and N. Yokoya, "Web-based telepresence system using omnidirectional video streams," *In Proc. of 5th Pacific Rim Conf. on Multimedia*, Vol. 3, pp. 45–52, 2004. (第2章に関連)

## 研究会・シンポジウム

1. 石川 智也, 山澤 一誠, 横矢 直和, "全方位ビデオ映像の実時間無線配信を利用した位置情報システム", 情報処理学会 研究報告 2008-CVIM-161, pp. 333–334, 2008. (第2章に関連)
2. 石川 智也, 山澤 一誠, 横矢 直和, "複数視点全方位映像のマルチキャストによる高スケーラビリティ自由視点テレプレゼンス", 3次元画像コンファレンス 2007, pp. 63–66, 2007. (第3章に関連)
3. 大野 純佳, 石川 智也, 山澤 一誠, 横矢 直和, "複数のPTZカメラを用いた保育施設における複数ユーザの要求に応じた映像提示", 電子情報通信学会 技術研究報告 PRMU2006-184, 2007.
4. 石川 智也, 山澤 一誠, 横矢 直和, "複数の全方位ライブ映像からの自由視点テレプレゼンス", 第2回デジタルコンテンツシンポジウム講演予稿集, No. 3-1, 2006. (第3章に関連)

5. 石川 智也, 山澤 一誠, 横矢 直和, ”複数の全方位動画像からの自由視点画像生成”, 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 1396–1403, 2005. (第3章に関連)
6. 石川 智也, 山澤 一誠, 佐藤 智和, 中村 豊, 藤川 和利, 砂原 秀樹, 横矢 直和, ”全方位カメラとWebブラウザやHMDによるネットワークテレプレゼンス”, 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 1606–1607, 2005. (第2章に関連)
7. 石川 智也, 山澤 一誠, 佐藤 智和, 中村 豊, 藤川 和利, 砂原 秀樹, 横矢 直和, ”全方位カメラとWebブラウザを用いたネットワークテレプレゼンス”, 第1回デジタルコンテンツシンポジウム講演予稿集, No. S6-1, 2005. (第2章に関連)
8. 石川 智也, 山澤 一誠, 横矢 直和, ”MorphingとVisual Hullを用いた動的環境での実時間自由視点画像生成”, 電子情報通信学会 技術研究報告 PRMU2004-136, 2004. (第3章に関連)
9. 石川 智也, 山澤 一誠, 佐藤 智和, 池田 聖, 中村 豊, 藤川 和利, 砂原 秀樹, 横矢 直和, ”Webブラウザと全方位動画像を用いたテレプレゼンスシステム 車載全方位カメラや全方位型マルチカメラシステムを用いた実装”, 画像の認識・理解シンポジウム (MIRU2004) 講演論文集, Vol. I, pp. 127–132, 2004. (第2章に関連)
10. 山澤 一誠, 石川 智也, 佐藤 智和, 中村 豊, 藤川 和利, 砂原 秀樹, 横矢 直和, ”車載全方位カメラと無線ネットワークを用いた実時間テレプレゼンス”, 情報処理学会 研究報告 CVIM142-9, 2004. (第2章に関連)

## 全国大会

1. 大野 純佳, 石川 智也, 山澤 一誠, 横矢 直和, ”複数のPTZカメラによる複数人の要求に応じた映像提示”, 電気関係学会関西支部連合大会講演論文集, No. G15-2, 2006.



2. 石川 智也, 山澤 一誠, 横矢 直和, "View Morphing と Visual Hull を用いた自由視点画像生成", 情報科学技術フォーラム (FIT) 一般講演論文集, Vol. 3, pp. 127-128, 2004. (第 3 章に関連)
3. 石川 智也, 山澤 一誠, 池田 聖, 横矢 直和, "Web ブラウザと全方位動画像を用いたテレプレゼンスシステム", 2004 年電子情報通信学会総合大会講演論文集, No. D-12-160, 2004. (第 2 章に関連)

## 解説

1. 石川 智也, 山澤 一誠, 横矢 直和, "自由視点テレプレゼンスシステム -多地点全方位ライブ映像による自由視点画像の生成と提示-", 画像ラボ, Vol. 18, No. 3, pp. 8-13, 2007. (第 3 章に関連)

## 表彰

1. 2007 年 2 月, IEEE 関西支部, 2006 年度 IEEE 関西支部学生研究奨励賞
2. 2005 年 6 月, 映像情報メディア学会, 第 1 回デジタルコンテンツシンポジウム船井賞

## マスメディア報道

1. "過去と未来が共存する町 生駒市 -生駒の未来力-", K-CAT (ケーブルテレビ), 2008 年 1 月 4 日.
2. "仮想世界「面白い」生駒でフェス", 産経新聞, 2006 年 4 月 25 日.
3. "高山サイエスタウンフェスティバル「学研都市の魅力を知って」", 奈良新聞, 2006 年 4 月 23 日.
4. "ロボット公道実験 先端科技大が近畿初 移動中の映像配信も", 産経新聞, 2005 年 11 月 12 日.