

NAIST-IS-MT0351012

修士論文

全方位動画像を用いたテレプレゼンスの ネットワーク拡張と自由視点画像生成に関する研究

石川 智也

2005年3月10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
修士(工学) 授与の要件として提出した修士論文である。

石川 智也

審査委員：

横矢 直和 教授

砂原 秀樹 教授

山澤 一誠 助教授

全方位動画像を用いたテレプレゼンスの ネットワーク拡張と自由視点画像生成に関する研究*

石川 智也

内容梗概

近年，現実世界を仮想世界に取り込み，時間的・空間的に隔たった場所の事象を体験する技術であるテレプレゼンスの研究が盛んに行われている．テレプレゼンス技術は現実環境の画像を使用するため3DCGで構築した仮想環境を提示するよりもリアリティが高く，エンターテインメントや遠隔監視，遠隔会議等様々な分野への応用が可能である．

テレプレゼンス技術を応用したシステムに要求される機能として，「広範囲な現実環境の撮影」「インタラクティブな画像提示」「ライブ動画像の処理と提示」「複数の利用者に対する画像提示」があげられる．仮想環境を利用者に提示する場合，任意の位置・方向の画像を利用者の動作に合わせて提示することにより，利用者にあたかも仮想環境中に居るかのような感覚を与えることができる．また，ライブ動画像を処理することにより，実時間性の求められる遠隔監視やTVのライブ中継のようなサービスが可能となる．さらに複数の利用者が同時にシステムを利用可能であるならば，TV放送のような多視聴者に対するサービスも可能となる．これらの条件を満たすことで複数の利用者に高臨場感のテレプレゼンスが可能となる．しかし，従来のシステムでは，これらの機能を全て満たすには到っていない．

そこで，本研究では以上の機能を全て満たすテレプレゼンスシステムを実現することを目標とする．そのために，まず二つのシステムを実現する．一つは，全

* 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 修士論文, NAIST-IS-MT0351012, 2005年3月10日.

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システムである。ネットワークテレプレゼンスシステムは、全方位カメラから取得した全方位動画をネットワーク上でマルチキャスト配信し、その動画を Web ブラウザと連携して動作するビューアにより受信することで、多地点の利用者が全方位動画を容易に鑑賞することが可能である。もう一つは、動的実環境を自由な視線方向・視点位置でテレプレゼンス可能な自由視点テレプレゼンスシステムである。自由視点テレプレゼンスシステムは、環境中に複数の全方位カメラを配置し、それらのカメラからの全方位動画を静的領域と動的領域に分離し、静的領域には Morphing、動的領域には Visual Hull を用いることによってそれぞれの領域の自由視点画像を生成する。そして、それらを合成することによって最終的な自由視点画像として提示する。

ネットワークテレプレゼンスの実験では、あらかじめ撮影しておいた高解像度の動画をサーバに格納しておき、その動画を LAN 内の PC で鑑賞した。さらに、屋外の車載全方位カメラの映像をマルチキャスト配信し、その映像を屋内の PC で鑑賞した。提案システムにより多地点の利用者が同時に全方位画像を鑑賞できることを確認した。一方、自由視点テレプレゼンスの実験では、環境中に配置した 3 つの全方位カメラのライブ動画を入力とし、実時間で自由視点画像を生成・提示し、本手法の有効性を確認した。

キーワード

全方位動画、テレプレゼンス、マルチキャスト、Web ブラウザ、自由視点画像生成、Morphing、Visual Hull

Networked Telepresence and Novel View Generation Using Omnidirectional Videos*

Tomoya Ishikawa

Abstract

Recently, telepresence which allows us to experience a remote site through a virtualized real world has been investigated. The telepresence technology which uses real scene images provides us with the sense of rich presence. Thus, it can be widely applied to a number of fields such as entertainment, remote surveillance and video conference.

Telepresence system is required to capture wide range of real world, to change user's view interactively, to process live video, and to present images to many users. When users see a virtualized real world, the users feel rich presence of the world by presenting images of user's view direction and position. Live video processing enables telepresence system to provide real-time services like remote surveillance. Further, presenting images to many users enables telepresence system to provide services like TV broadcasting. The functions enable telepresence system to provide rich presence to many users. However, conventional systems do not implement all of the functions.

The objective of this study is realization of immersive telepresence system which have the functions. First of all, I realize two systems for that. The one is networked telepresence system using omni-directional cameras and web browsers. The networked telepresence enables users in different places to look around a

* Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0351012, March 10, 2005.

virtualized real world easily in network environments. The system is a server-client system. The server transfers omni-directional video streams using a multi-cast protocol and users can see the video easily through omni-directional video viewers using web browsers at different sites. The another is novel view-point telepresence system. The novel view-point telepresence system capture omni-directional images from omni-directional cameras and the images are divided into static and dynamic regions. Novel view images are generated by morphing of static regions and by visual hull of dynamic regions. A user sees a image which is synthesized from two novel view images.

In the experiment of networked telepresence system, two types of systems have been realized: store-and-playback system and real-time system. In the former system, stored high-resolution omni-directional videos are interactively presented via network, while the latter enables users to observe omni-directional live videos transferred from a moving car. The captured live video is transferred to user's PC through wireless and wired network. In the experiment of novel view-point telepresence system, we have implemented a prototype of novel view generation in real-time from omni-directional live video streams. The feasibility of the proposed systems and methods has been confirmed with both experiments.

Keywords:

omni-directional video, telepresence, multi-cast, web browser, novel view synthesis, morphing, visual hull

目次

1. はじめに	1
1.1 テレプレゼンスに関する背景	2
1.2 固定視点テレプレゼンスの従来研究	4
1.3 自由視点テレプレゼンスの従来研究	7
1.3.1 モデルベースドレンダリング	7
1.3.2 イメージベースドレンダリング	10
1.3.3 ハイブリッド手法	14
1.4 本研究の方針と位置づけ	16
2. テレプレゼンスシステムのネットワーク拡張	20
2.1 ネットワークテレプレゼンスシステムの概要	20
2.2 全方位カメラ	21
2.2.1 全方位カメラ HyperOmni Vision	21
2.2.2 全方位マルチカメラシステム Ladybug	23
2.3 全方位動画像コンテンツ	24
2.4 Web ブラウザを用いた全方位動画像ビューア	25
2.5 実験	27
2.5.1 高解像度の蓄積型動画像を用いた実験	27
2.5.2 車載全方位カメラとマルチキャストを用いたライブ型動画 像による実験	30
2.6 考察	34
3. 全方位画像を用いた動的環境における自由視点画像生成	35
3.1 静的領域と動的領域を分離したハイブリッド手法の概要	35
3.2 全方位画像カメラによる画像取得	35
3.3 静的領域と動的領域の分離	36
3.3.1 背景更新を行う背景差分処理による動的領域の検出	37
3.3.2 動物体の方位角情報の取得	38

3.4	静的領域に対する Morphing を用いた自由視点画像生成	38
3.5	動的領域に対する Visual Hull を用いた自由視点画像生成	39
3.5.1	Visual Hull の算出処理	40
3.5.2	動物体の位置推定による処理の削減	40
3.6	静的領域と動的領域の自由視点画像の合成	42
3.7	実験	42
3.7.1	システム構成	44
3.7.2	結果	44
3.8	考察	48
4.	まとめと今後の課題	49
	謝辞	51
	参考文献	52

目 次

1	森田らのシステム構成 [33]	5
2	池田らのテレプレゼンスシステムの外観 [7]	6
3	PanoVi のシステム構成 [34]	7
4	レーザーレンジファインダ	8
5	3DRoom と三次元形状復元した動物体 [12]	9
6	View Morphing [18]	11
7	Image-Based Visual Hull [24]	12
8	稲本らの自由視点画像生成手法の概要 [25]	13
9	稲本らの手法により生成した自由視点画像 [25]	13
10	ビルボーディング技術を用いた自由視点画像生成 [27]	14
11	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 [28]	15
12	環境の 2.5 次元モデルと自由視点画像 [29]	16
13	全方位自由視点画像と平面透視投影画像 [30]	16
14	ネットワークテレプレゼンスシステムの概要	21
15	HyperOmni Vision	22
16	全方位画像と平面透視投影画像	23
17	全方位マルチカメラシステム Ladybug	24
18	池田らの手法 [7] により生成された全方位画像	24
19	全方位動画像ビューアの構造	26
20	鑑賞用 PC の Web ブラウザ画面	28
21	ジャイロセンサを取り付けた HMD	29
22	ライブ型動画像による実験のシステム構成	31
23	画像取得車両の外観	32
24	画像取得車両の内部	32
25	利用者それぞれの全方位動画像ビューアの画面例	33
26	提案手法の処理の流れ	36
27	極座標変換によるヒストグラムの作成	38
28	対応点の自由視点画像への投影	39

29	共通領域の有無の判定	41
30	推定される存在領域	43
31	直線の媒介変数表示	43
32	自由視点画像生成の実験環境	45
33	プロトタイプシステムの構成	45
34	実験環境の様子と生成された自由視点画像	47

表 目 次

1	蓄積型動画鑑賞用 PC の詳細	28
2	画像取得車両のシステム構成	32

1. はじめに

近年，遠隔地の現実環境を仮想化し，時間的・空間的に隔たった場所の事象をあたかもその場に居るかのような感覚で体験する技術であるテレプレゼンス [1] の研究が盛んに行われている．テレプレゼンス技術において，利用者に体験させる事象の情報として，画像・音声・触覚・においなどがある．このうち画像は遠隔地の様子を理解する上で最も重要な情報である．テレプレゼンスにおいて，利用者に提示する画像は，遠隔地の現実環境を撮影したものが用いられるため，3DCG によって構築された仮想環境を提示するよりもリアリティが高い．また，テレプレゼンス技術は，エンターテインメントや遠隔監視，遠隔会議など様々な分野への応用が可能である．

テレプレゼンス技術を応用したシステムに要求される機能として，以下のようなものがあげられる．

- 広範囲な現実環境を撮影
- インタラクティブな画像提示
- ライブ動画像の処理と提示
- 複数の利用者に対する画像提示

仮想環境を利用者に提示する場合，任意の位置・方向の画像を利用者の動作に合わせて提示することにより，利用者にあたかも仮想環境中に居るかのような感覚を与えることができる．また，ライブ動画像を処理することにより，実時間性の求められる遠隔監視や TV 中継のようなサービスが可能となる．さらに複数の利用者が同時にシステムを利用可能であるならば，TV 放送のような多視聴者に対するサービスも可能である．これらの条件を満たすことで複数の利用者に臨場感の高いテレプレゼンスが可能となる．

本研究では，以上のような機能を持つテレプレゼンスシステムの実現を目標とする．以下本章ではテレプレゼンスに関する背景とその従来研究を概観し，本研究の方針と位置づけについて述べる．

1.1 テレプレゼンスに関する背景

テレプレゼンスを実現する簡単な手段として、CCD カメラ等の画像センサにより取得した画像を遠隔地に送り、その画像を提示することが考えられる。しかし、一般的なカメラの画角は 35 ~ 60 度程度であり、広範囲の現実環境の撮影は困難である。この問題を解決する手法としてカメラを方向が制御できる回転台に設置し、遠隔地からカメラの方向を制御することで広範囲の現実環境の撮影することが考えられる。この手法では、広範囲の撮影とインタラクティブな画像提示が可能であるが、カメラの方向を変更するには回転台を動かす必要があり、利用者が視線方向の変更を指示してその視線方向の画像が撮影されるまでに機械的な時間遅延が発生する。また、視線方向の変更指示を回転台に伝送する際にも時間遅延が発生する。これらの時間遅延により、利用者は仮想環境への没入感を損なってしまう。さらに、この手法ではカメラと利用者は 1 対 1 となり、複数の利用者がテレプレゼンスを行うのは困難である。

回転台を使用せずに広範囲を撮影するために、全方位カメラを使用する手法がよく用いられる。全方位カメラとは、周囲 360 度を一度に撮影可能なカメラを指し、魚眼レンズ [2] や曲面ミラー [3, 4] によって一般的なカメラで広範囲を撮影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範囲を撮影する全方位マルチカメラシステム [5, 6, 7] 等がある。これら全方位カメラで撮影された全方位画像から利用者の視線方向の画像を生成し提示することで機械的な時間遅延のない視線方向の変更が可能である。さらに、一つの全方位カメラから撮影された画像を用いて複数の利用者が同時に異なる視線方向の画像を見ることができるといった特徴を持つ。しかし、魚眼レンズや曲面ミラーによる全方位カメラは周囲 360 度の情景を写すため角度分解能は一般的なカメラよりも低くなってしまふ。全方位マルチカメラシステムの角度分解能は一般的なカメラとほぼ同等であるが、画像サイズが大きく、実時間での符号化・伝送を行うのは困難である。広範囲の環境を撮影する目的においては、魚眼レンズやミラーによる全方位カメラと全方位マルチカメラシステムを用途によって適切に選択すれば、回転台にカメラを設置し視線方向を変更するシステムに比べ有用である。

以上では、ある視点位置において広範囲の環境を撮影し、その画像を用いるテ

レプレゼンスについて述べた。このようなある撮影位置にカメラを配置し、その撮影位置における画像を提示するテレプレゼンスを固定視点テレプレゼンスと呼ぶ。一方、撮影位置以外の視点位置における画像を生成することで利用者の視点移動を可能にするテレプレゼンスが盛んに研究されており、これを自由視点テレプレゼンスと呼ぶ。テレプレゼンスにおいて、インタラクティブに利用者の視線方向を変更することと同様に視点位置の変更も行えることがよりテレプレゼンスを臨場感豊かにする。現実環境を撮影した位置とは異なる視点位置における画像を生成する技術は自由視点画像生成技術と呼ばれ、その手法は大きく3つに分類することができる。環境を撮影した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からの画像を生成するモデルベースドレンダリング (MBR:Model-Based Rendering) [8, 9, 10, 11, 12, 13, 14, 15]、環境の3次元形状を復元せずに画像から直接自由な視点からの画像を生成するイメージベースドレンダリング (IBR:Image-Based Rendering) [16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]、モデルベースドレンダリングとイメージベースドレンダリングを組み合わせたハイブリッド手法 [27, 28, 29, 30] の3つである。

モデルベースドレンダリング手法は、環境の3次元形状を復元する。環境の形状を持っているため、利用者の視点移動の自由度や視点移動に伴う奥行き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を移動させることが容易に行える。しかし、複雑な形状を持つ環境や物体はモデル化が困難である。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量が必要となり、画像の取得から自由視点画像の提示までを実時間で行うことは困難である。

イメージベースドレンダリング手法は、自由視点画像生成のための計算量が環境の3次元形状の複雑さに依存せず一定である。一般的にこの手法はモデルベースドレンダリング手法に比べ高速に処理が可能であり、画像の取得から自由視点画像の提示までを実時間で行うことが可能である。しかし、多くのイメージベースドレンダリング手法では複数の地点で撮影された画像間の対応関係を必要とすることや、密に画像を取得する必要がある。

モデルベースドレンダリング手法とイメージベースドレンダリング手法を組み

合わせることで、それぞれの手法の特徴を併せ持つハイブリッド手法では、物体の三次元形状を正確には復元せず、利用者への画像の見せ方を工夫することでモデルベースレンダリング手法と同等の自由視点画像を生成する。

1.2 固定視点テレプレゼンスの従来研究

前節では固定視点テレプレゼンスによく用いられる回転台にカメラを設置し視線方向を変更可能とするシステムと全方位カメラにより撮影される画像を用いて視線方向を変更可能とするシステムについて述べた。本研究では、目標とするテレプレゼンスシステムを実現するには全方位カメラや全方位画像を用いたシステムが有効であると考え、それらに焦点を当て従来研究を概観する。

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にするシステムに QuickTimeVR[31] がある。QuickTimeVR はカメラを回転して得られたパノラマ画像を円筒面に投影し、その円筒内から任意の方向の円筒面を見ることにより周囲 360 度を見回すことを可能にしている。さらに、Web ブラウザ上で動作する等、利用者が簡単にテレプレゼンスを行えるシステムとなっている。しかし、QuickTimeVR で扱える画像は静止画のみであり、全方位動画像を用いて動的環境を見回すことはできない。

尾上らのシステム [32] は QuickTimeVR の問題であった動的環境の仮想化を全方位カメラを用いて解決した。このシステムは一度に周囲 360 度を撮影可能な全方位カメラによって全方位動画像を取得し、HMD と磁気トラッカを組み合わせて利用者の視線方向の変化に合わせた画像提示を行う没入感の高いシステムである。しかし、全方位カメラと画像提示用 PC は直接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、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 [33] は、尾上らのシステムをネットワークを介した映像伝送に改良し、さらにそのシステムを応用した遠隔監視システムを提案した。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 Digital Video Transport System (DVTS) を用いてネットワークを通し、伝送する。DVTS は高画質の動画像を少ない遅延で伝送できる代わりに広帯域幅を必要とする。また Peer to Peer 接続に

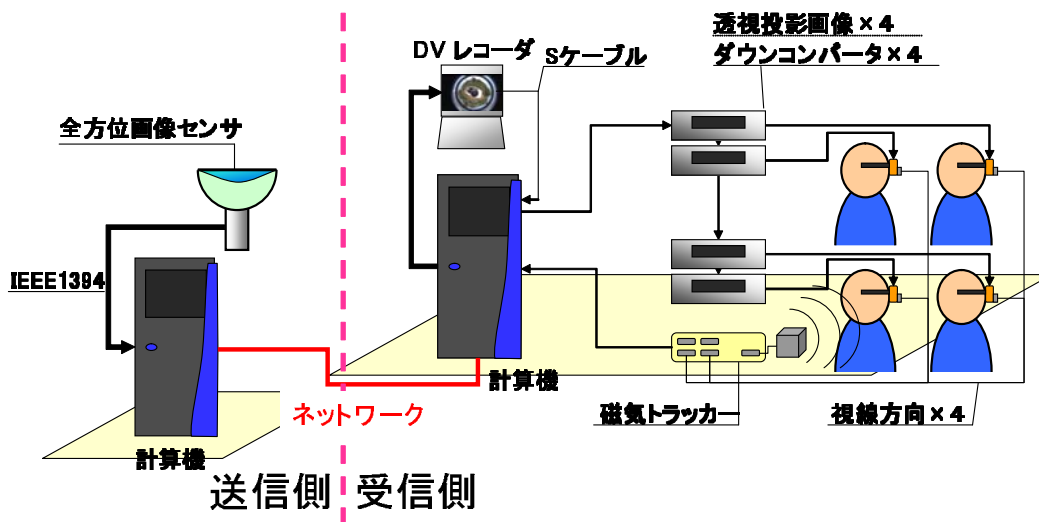


図 1 森田らのシステム構成 [33]

よって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 1 参照）．画像の提示には尾上らのシステム同様，HMD と磁気トラッカと組み合わせたものを使用している．このシステムでは映像の伝送が Peer to Peer であるため，一つの撮影地点からの映像を多地点の遠隔地で見ることが困難である．また，DV カメラによって周囲 360 度の画像を取得するため実際に利用者が見る画像は解像度が低くなっている．

魚眼レンズやミラーによる全方位カメラでは，全方位画像の解像度が低いことが問題となる．そこで全方位マルチカメラシステムにより周囲を撮影するシステムが考えられる．その一つに全方位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，その画像を球面ディスプレイに投影する池田らのシステム [7] がある．このシステムでは，予め全方位カメラシステムを光学的・幾何学的に高精度なキャリブレーションをしておき，その情報を基に違和感の少ない高解像度な全方位画像を生成する．そして，その画像を球面ディスプレイに投影する．システムの利用者はゲームパッドによって視線方向を自由に変更することができる．図 2 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．しかし，このシステムで



図 2 池田らのテレプレゼンスシステムの外観 [7]

は、環境撮影中の画像をディスクアレイに保持し、撮影後に画像のモザイクングを行う必要があるため、ライブ動画像を用いてテレプレゼンスを行うのは困難である。

全方位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て遠隔地の動的環境を見回すことが可能なシステムに PanoVi[34] が提案されている。PanoVi のシステム構成を図 3 に示す。PanoVi は広角な 4 台のカメラを 90 度毎に配置した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、ネットワークを通して遠隔地へ伝送する。ただし、この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 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。全方位動画像の取得から伝送・提示までをライブで行うことができるシステムで

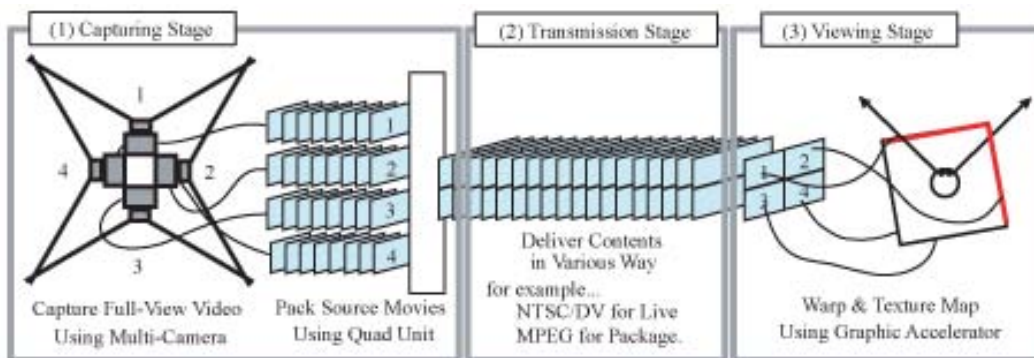


図 3 PanoVi のシステム構成 [34]

あるが、クライアント PC 側で 4 つのカメラから得られた画像をモザイクするためクライアント PC に負荷がかかることや、実時間でのモザイクでは画像間の色調の違いなどに十分に対応しきれないことによる画像間の継ぎ目の表れ等の問題がある。また、映像をユニキャストで伝送するため、利用者が増えると伝送に必要な帯域が増加するといった問題がある。

1.3 自由視点テレプレゼンスの従来研究

仮想環境中を自由にウォークスルーすることを可能にする自由視点画像生成技術についての研究も盛んに行われている。この自由視点画像生成技術はモデルベースレンダリング手法とイメージベースレンダリング手法とそれらのハイブリッド手法の 3 つに分けることができる。

1.3.1 モデルベースレンダリング

モデルベースレンダリング手法は現実環境中の物体の形状を取得し、仮想環境中にそれらを配置することにより自由な視点移動を可能とする。物体の 3 次元モデルを持つため、現実とは異なる位置に物体を移動したり、3DCG によって生成した仮想物体を加えることも可能である。この手法において、現実環境中の物体の形状を取得する方法として、レーザーレンジファインダを使用して物体の形



Leica 社製 TCR1105XR



Riegl 社製 LMS-Z360

図 4 レーザーレンジファインダ

状を計測する方法 [8, 9] やカメラで撮影した画像からコンピュータビジョン技術により物体の形状を復元する方法 [10, 11, 12, 13, 14, 15] がある。

レーザーレンジファインダ (図 4 参照) は、レーザを照射し物体までの距離を計測する。そのため、物体の形状を高精度に取得可能であるが、形状の情報を密に取得するほど計測に時間がかかる。また、広範囲の環境の形状を取得するには多地点で計測を行う必要があり、それに伴い多地点で計測されたモデルを一つに統合する処理が必要となる。これらの特徴から、レーザーレンジファインダを使用して物体の形状を計測する方法は静的環境の仮想化には適しているが、動的環境の仮想化には適していない。

一方、カメラで撮影した画像からコンピュータビジョン技術により物体の形状を復元する方法は、多地点で撮影された画像における物体のずれから三角測量の原理に基づき物体までの距離を推定する手法や多視点で撮影される物体のシルエットから物体の近似的な 3 次元形状を復元する手法などがある。

三角測量の原理を用いて距離を推定する手法は、画像間で対応点を必要とする。その対応点が密に取得できれば物体の形状をより正確に表現できる。しかし、画

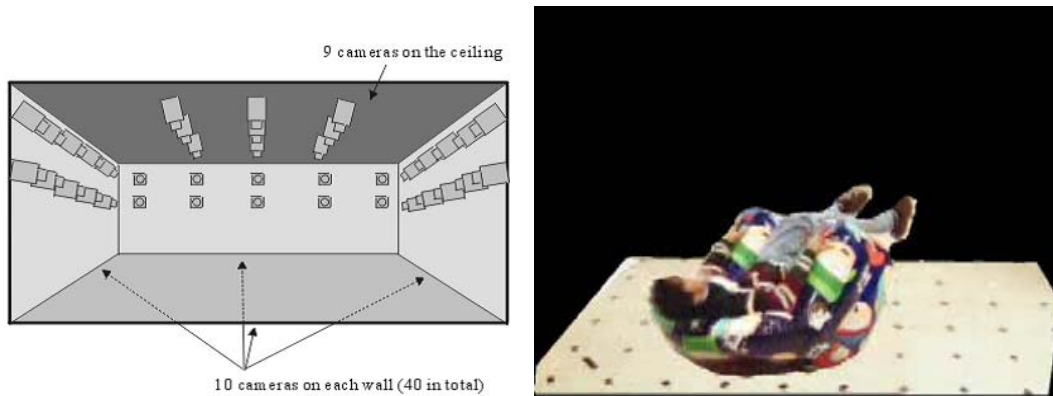


図 5 3DRoom と三次元形状復元した動物体 [12]

像間の対応点を何の情報も無く正確に与えることは非常に困難である．佐藤らの手法 [10] は，ビデオカメラで三次元位置が既知の点と自然特徴点を追跡していくことにより画像間の対応を与え，三次元形状を復元する．この手法のようにビデオシーケンスにおいて，画像上の点を追跡することで画像間の対応点を自動で与えることが可能であるが，複数のビデオフレームから対応点の三次元位置を推定するため，撮影環境中の動的物体などの形状を復元することは困難である．

撮影時のカメラの位置・姿勢が既知であれば，マルチベースラインステレオ [11] によって画像間の対応点を与えることが可能である．斎藤ら [12] や Narayanan ら [13] は位置・姿勢が既知のカメラを複数配置した 3DRoom と呼ばれる室内環境（図 5 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三次元復元を行った．しかし，三次元形状復元に掛かる計算処理コストが高く実時間での処理は困難であることや室内の限られた範囲の物体でなければ三次元形状を得られない．よって，ライブ動画像の処理や提示，広範囲の環境の再現は行えない．

以上のような多地点の画像の対応点から距離を推定することなく，撮影対象の三次元形状を推定する手法も提案されている．視体積交差法 [14] は多地点で得られた撮影対象のシルエット画像から対象物体の近似的三次元形状を得る手法である．この手法は撮影空間をボクセルと呼ばれる立方体で格子状に表現したボクセル空間内で，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) を求め、それをボクセルの集合で表現する。この 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を用いることによって物体の三次元形状に近づく性質がある。この手法はシルエット画像が得られた時のカメラの位置・姿勢は既知である必要があるが、三角測量に基づく手法と異なり、対応点探索の問題が無い。また、処理速度はボクセル空間の解像度に依存するが、ステレオ法による三次元形状復元と比較すると高速である。しかし、この手法は斎藤ら [12] や Narayanan ら [13] の手法同様に、多地点から対象物体を撮影する必要があり、撮影範囲が限定されやすい。さらに、三次元形状を復元する物体は撮影画像上の小領域でなければならず、広範囲の環境の三次元形状復元は非常に困難である。

1.3.2 イメージベースドレンダリング

環境の明確な三次元形状を用いるモデルベースドレンダリング手法に対し、イメージベースドレンダリング手法は三次元形状を用いることなく自由視点画像生成を行う。環境の三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一定である。また、一般的にモデルベースドレンダリング手法に比べ高速に処理が可能である。

イメージベースドレンダリングの自由視点画像生成手法として、代表的な手法に Warping や Morphing による手法 [16, 17, 18, 19] が提案されている。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の位置がどこに現れるかを求め、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ディングすることによって決めるというものである (図6参照)。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、モデルベースドレンダリング手法同様、対応点を自動的にかつ正確に与えることは困難である。山崎らの手法 [16] は、入力画像に関する知識なしに自然な対応点を与えることを可能にしている。しかし、対応を求めるのに非常に処理コストがかかるため動的環境の自由視点画像生成に適用するのは困難である。

対応点を必要としないイメージベースドレンダリング手法に、空間中を伝播する

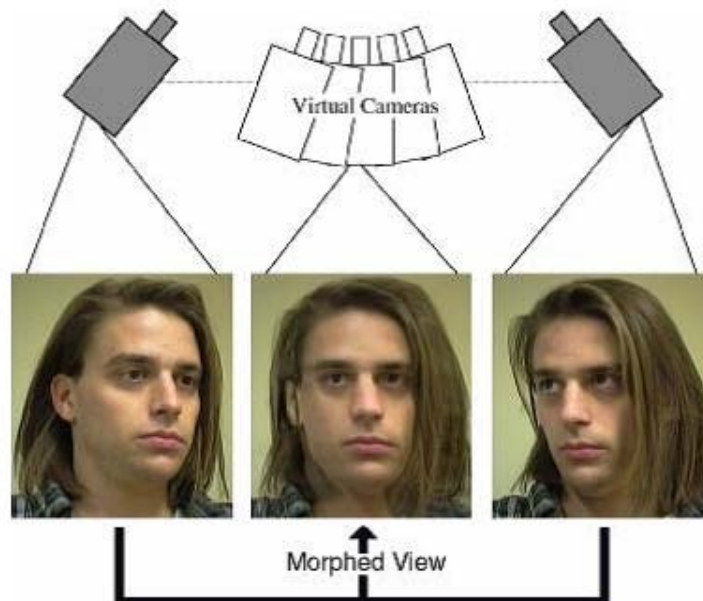


図 6 View Morphing [18]

光線を記録し，自由視点位置において画像平面を考え，その平面を通過する光線を画像化することで自由視点画像生成を行う手法が提案されている [20, 21, 22, 23] . この手法は対応点を必要としないが，画像を撮影した時のカメラの位置・姿勢情報が必要である．また，空間中の光線情報を収集するために環境を密に撮影する必要があるため，広範囲の環境には適用困難である．

Matusik ら [24] は物体の明確な三次元形状を復元することなく，自由視点における物体の見えを計算する手法を提案している．この手法は入力シルエット画像とその画像が得られたカメラの位置・姿勢によるエピポラ方程式を基に，Visual Hull を計算し自由視点画像生成を行っている（図 7 参照）．さらに 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．この手法も対応点を必要としないが，視体積交差法同様に周辺の広範囲の環境の自由視点画像を生成することは困難である．図 7 右の人物以外の周辺環境は予め作成された 3DCG である．

稲本ら [25] は，複数のイメージベースドレンダリング手法を組み合わせること

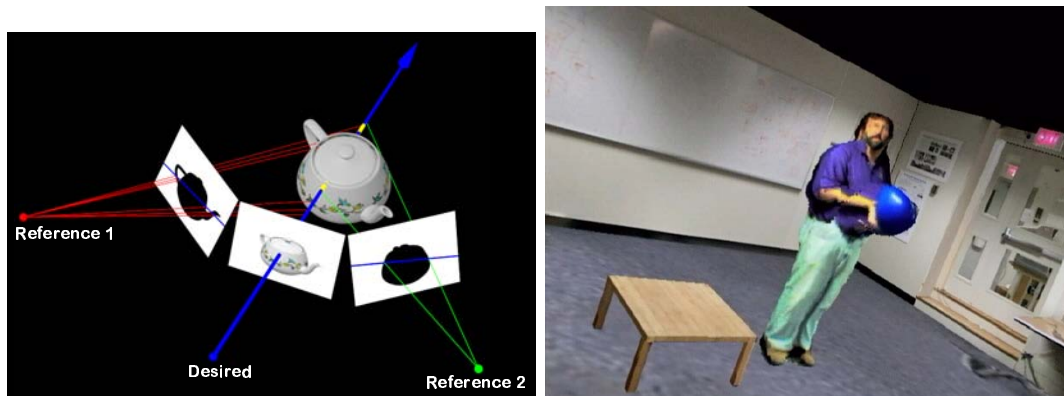


図 7 Image-Based Visual Hull [24]

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 8 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ング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ビポラ方程式を利用した対応付けを行い 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 9 参照）．以上のように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が，屋内環境やその他の静的領域を少ない平面と見なすことができない環境には適用困難である．

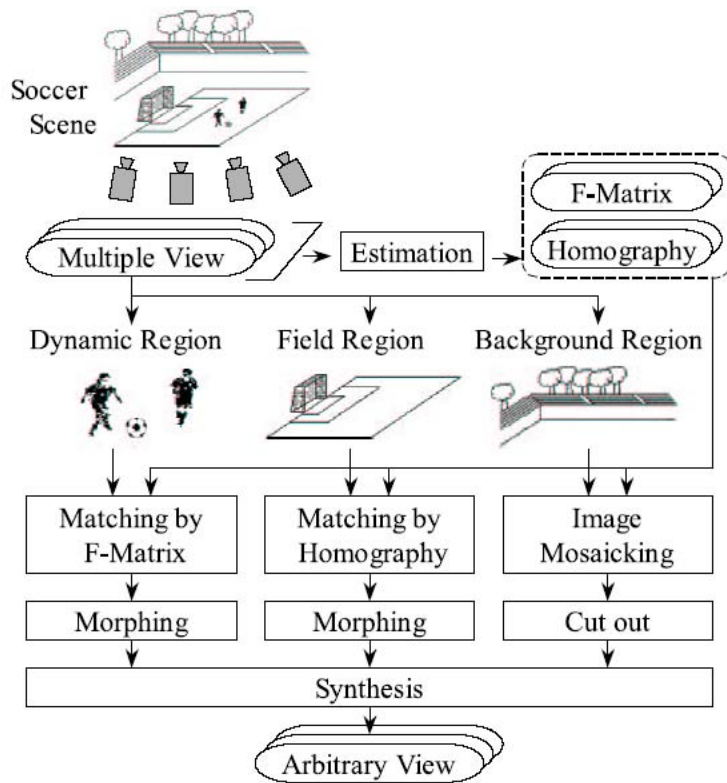


図 8 稲本らの自由視点画像生成手法の概要 [25]



図 9 稲本らの手法により生成した自由視点画像 [25]



図 10 ビルボーディング技術を用いた自由視点画像生成 [27]

1.3.3 ハイブリッド手法

モデルベースレンダリング手法とイメージベースレンダリング手法とを組み合わせたハイブリッド手法も提案されている。古山ら [27] は、サッカーシーンを自由視点で鑑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、その手法を利用したライブ動画像からサッカーを自由視点で鑑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。そこでこの手法は多視点で得られるサッカーシーンの画像からサッカープレイヤーの領域を検出し、プレイヤーの位置を求める。そして、そ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する。視点位置がサッカープレイヤーから離れていれば違和感を感じることはないが、視点位置がサッカープレイヤーに近づくと違和感を感じる。また、この手法を利用してネットワークを用いたライブ動画像の配信も実現しており、サッカーフィールドは予め 3DCG で作成されたものを配置している (図 10 参照)。

古山ら [27] と同様に、北原ら [28] は平面の集合によって近似的に対象物体の三

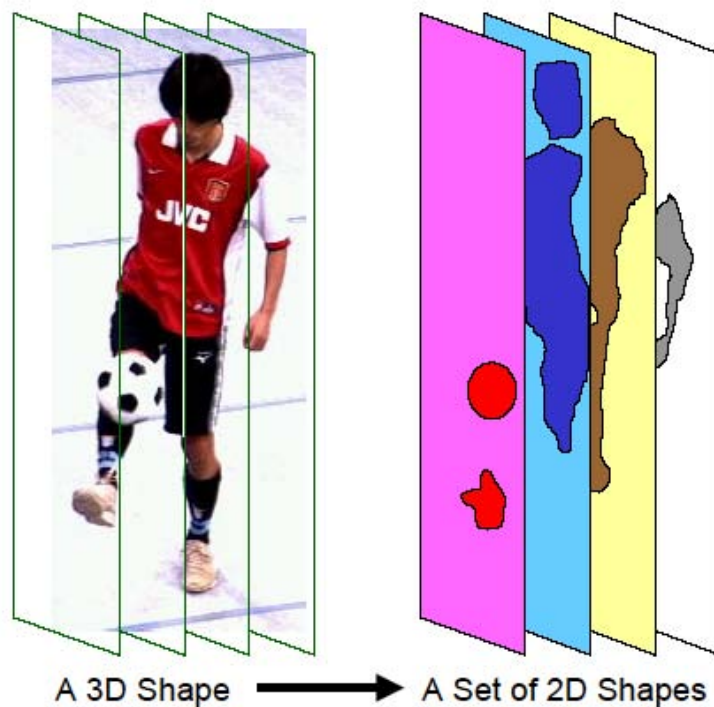


図 11 平面の集合によって近似的に三次元形状を復元する手法 [28]

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している．ボクセル等の方法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．さらにこの手法は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三次元形状の空間解像度を変化させ三次元形状復元の冗長性を減少させている．しかし，この手法も視体積交差法と同様に周辺環境を表現することは困難である．

島村ら [29] や富手ら [30] は周辺環境についても自由視点で見ることができるよう考慮した手法を提案している．島村らの手法 [29] 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環境を表現する（図 12 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．富手らの手法は

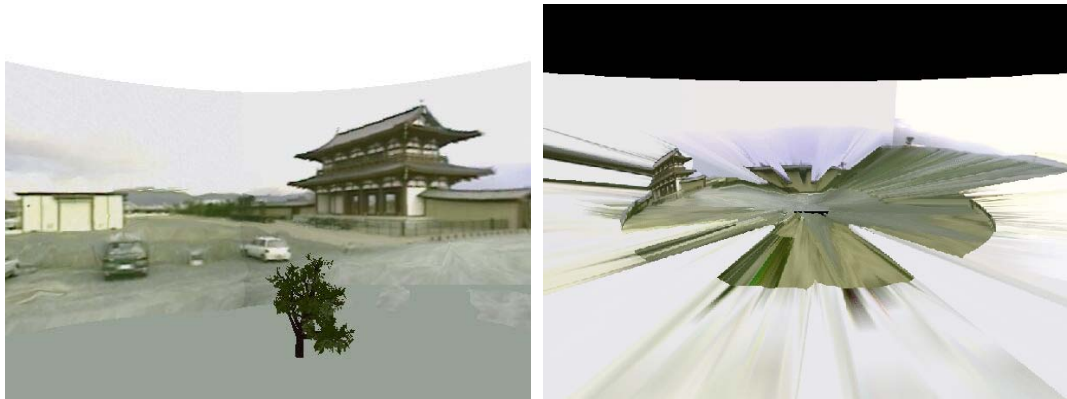


図 12 環境の 2.5 次元モデルと自由視点画像 [29]

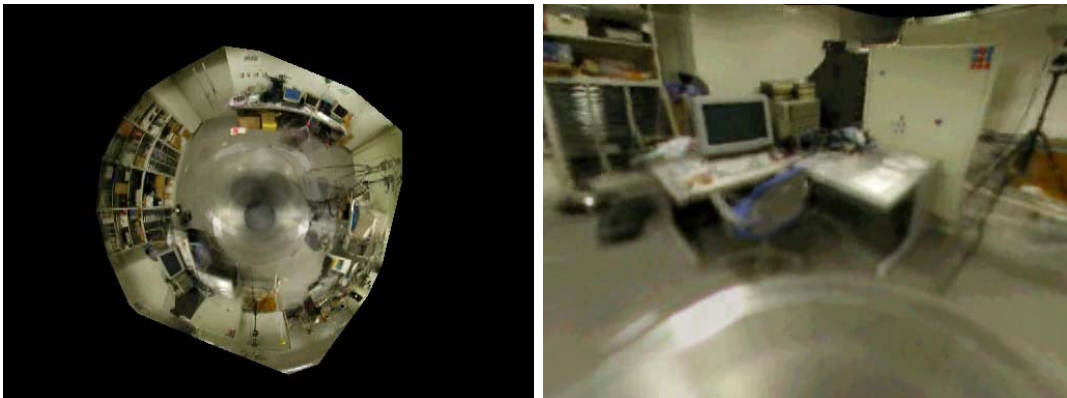


図 13 全方位自由視点画像と平面透視投影画像 [30]

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 Morphing を行う（図 13 参照）．その際，自由視点画像上の対応点の位置を対応点の三次元位置を基に計算する．この手法も対応点を必要とするため，動的環境に適用するのは困難である．

1.4 本研究の方針と位置づけ

本研究の目的は先に述べた以下の機能を持つテレプレゼンスシステムの実現である．

- 広範囲な現実環境を撮影
- インタラクティブな画像提示
- ライブ動画像の処理と提示
- 複数の利用者に対する画像提示

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、さらに本研究の位置づけを述べる。

広範囲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 [2, 3, 4, 5, 6, 7] を使用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ライブ動画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。

全方位カメラにより視線方向の自由度を得るが、自由な視点移動を可能にする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目標のテレプレゼンスシステムに適用可能かどうか考察する。

まず、モデルベースドレンダリング手法について考える。レーザーレンジファインダを用いる手法 [8, 9] 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。カメラで撮影された画像を用いる手法のうち、ステレオ法による距離推定を行う手法 [10, 11, 12, 13] は計算コストや対応点探索の問題から動的環境に適用は難しい。シルエット画像から物体の近似形状を得る手法 [14, 15] 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、目標とするテレプレゼンスシステム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。よって、モデルベースドレンダリング手法は本研究で目標とするテレプレゼンスシステムに適用困難である。

次に、イメージベースドレンダリング手法について考える。Warping や Morphing による手法は [17, 18, 19] 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 [16] 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 [20, 21, 22] は密に画像を取得する必要がある、広範囲の動的環境への適用は困難である。Matusik らの手法 [24] は、シルエット画像を基に自由視点画像を生成するため広範囲を再現することは困難である。稲本らの手法 [25] 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手法も目標とするテレプレゼンスシステムに適用困難である。

次に、ハイブリッド手法について考える。古山らの手法 [27] は視点位置と対象物体が離れている必要があり、システムの利用シーンが限定されすぎてしまう。北原らの手法 [28] は、広範囲の環境を再現することは困難である。島村らの手法 [29] 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 [30] は対応点を必要とするため動的環境に適用困難である。よって、目標とする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。

そこで、自由な視点移動を可能にするために、撮影環境中の静的領域と動的領域を分離して処理し、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。この事を考慮すると、撮影環境中の動的領域については Matusik ら [24] や北原ら [28] の手法を、静的領域については冨手らの手法 [30] 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る。本研究では、動的領域には Matusik らの手法、静的領域には冨手らの手法を用いることとした。

以上の考察より、本研究では目標のテレプレゼンスシステムを実現するために以下のアプローチをとる。

- 全方位カメラによる広範囲の環境の撮影
- マルチキャストプロトコルによる多地点への映像伝送

- 静的領域と動的領域に分離したハイブリッド手法による自由視点画像生成

本研究ではまず、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利用できるシステムと静的領域と動的領域に分離したハイブリッド手法による自由に視点移動が可能なシステムを実現する。そして、最終的にそれらのシステムを融合し、目標のテレプレゼンスを実現することを目指す。本論文では、それぞれのシステムの実現までを報告する。以下、2章では本研究で用いる全方位カメラの説明と従来のシステムをネットワーク拡張することにより複数人が同時かつ容易にテレプレゼンス可能なシステムについて述べる。3章では動的環境を自由な視点でテレプレゼンス可能なシステムについて述べる。そして、最後に4章で本論文のまとめと今後の課題について述べる。

2. テレプレゼンスシステムのネットワーク拡張

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利用できるシステムを提案する．さらに，Web ブラウザによる全方位動画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ようなシステムを構築する．

2.1 ネットワークテレプレゼンスシステムの概要

本ネットワークテレプレゼンスシステムの概要を図 14 に示す．本システムでは，全方位カメラにより全方位動画を撮影する．全方位動画はネットワーク越しのサーバや PC 内に格納されており，Web ブラウザにより起動されるビューアによってその全方位動画を取得する．あるいは，カメラからの映像をリアルタイムに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動画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，複数の利用者が配信された動画を受信してもネットワークの負荷は増加することはない．システムの利用者は 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動画コンテンツを簡単かつインタラクティブに鑑賞することができる．さらに，Head Mounted Display (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同様のビューアを用いて利用できる．

以下，2.2 節では提案システムで使用する全方位カメラについて述べる．2.3 節では，提案システムで扱う全方位動画の種類とその用途について述べる．2.4 節では，提案システムにおいて全方位動画を表示するために用いる全方位動画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実験について述べ，2.6 節でその考察を述べる．

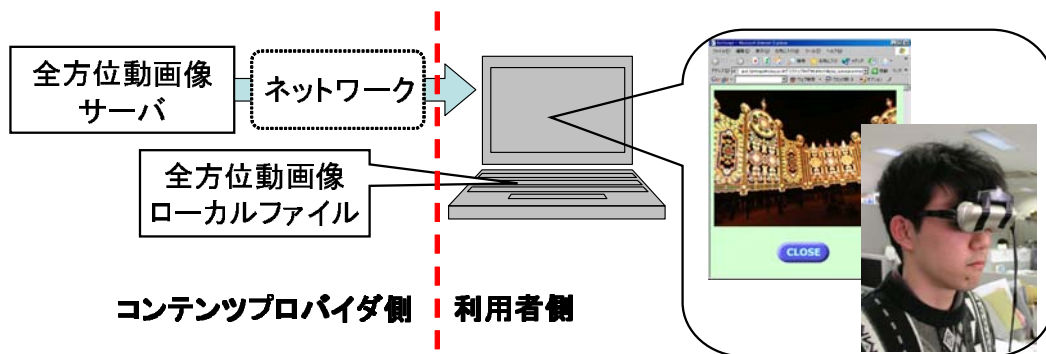


図 14 ネットワークテレプレゼンスシステムの概要

2.2 全方位カメラ

本節では、環境の撮影に用いる全方位カメラについて述べる。本研究では、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と複数のカメラを外向きに配置した全方位マルチカメラシステムを使用する。ミラーを用いた全方位カメラには HyperOmni Vision[4] を全方位マルチカメラシステムには Ladybug[7] を用いた。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。

2.2.1 全方位カメラ HyperOmni Vision

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。図 15 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 15 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)。

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

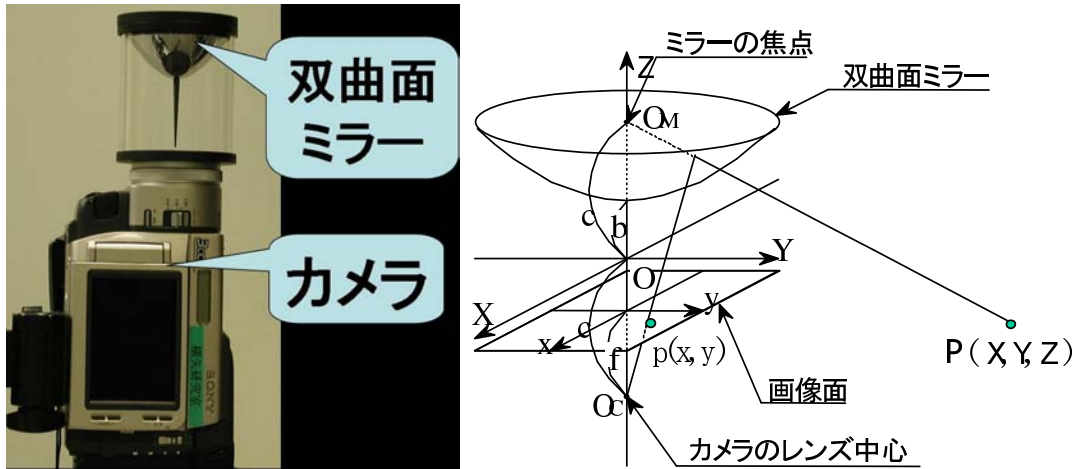


図 15 HyperOmni Vision

$$\text{レンズ中心 } O_C : (c = \sqrt{a^2 + b^2})$$

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。

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

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

式 (2) より入力全方位画像をミラー焦点 O_M から見た平面透視投影画像に変換することも可能である。

図 16 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



図 16 全方位画像と平面透視投影画像

し、ビデオカメラを用いることにより容易に全方位動画像が得られるという利点がある。

2.2.2 全方位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マルチカメラシステムである。図 17 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 17 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 17 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いる。一つの CCD カメラは横 768 ピクセル，縦 1024 ピクセル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 分間の動画像を蓄積できる。

本研究では，池田らの研究 [7] により生成された全方位動画像を用いる。図 18 に池田らの手法により生成された全方位画像を示す。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動画像を生成するため，ライブ動画像の撮影には向かないが上方向を含む広範囲の高解像度な画像が取得可能である。



図 17 全方位マルチカメラシステム Ladybug



図 18 池田らの手法 [7] により生成された全方位画像

2.3 全方位動画コンテンツ

全方位動画は、あらかじめ符号化し、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動画とライブ中継のようにリアルタイムに画像を符号化し配信するライブ型動画に分けられる。本システムではその両方の動画を扱うことが可能である。

蓄積型動画は、リアルタイムに画像を符号化し、伝送することが難しいような高解像度の全方位動画を予め符号化して PC 内に格納しておき、鑑賞するといった目的に用いる。蓄積型動画では利用者それぞれが、見たいシーンを見たい時刻から見るといったオンデマンドサービスが提供可能である。

ライブ型動画は、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。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に伝送し複数人が同時に鑑賞することが可能である。

2.4 Web ブラウザを用いた全方位動画ビューア

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 ページ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である。

図 19 に本システムで全方位動画を提示するソフトウェア（以下、全方位動画ビューア）の構造を示す。全方位動画ビューアはグラフィックスカードの機能を使用するため、ActiveX プログラムとして実装を行う。全方位動画ビューアは Web ブラウザにより起動され、全方位動画を平面透視投影画像に変換した後、Web ページ上にその画像を提示する。以下にその処理手順を述べる。

1. まず、Web ブラウザにより全方位動画ビューアが起動される。全方位動画ビューアは、受信する全方位動画のファイルパスやその全方位動画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 ページから読み取る。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鑑賞可能となる。

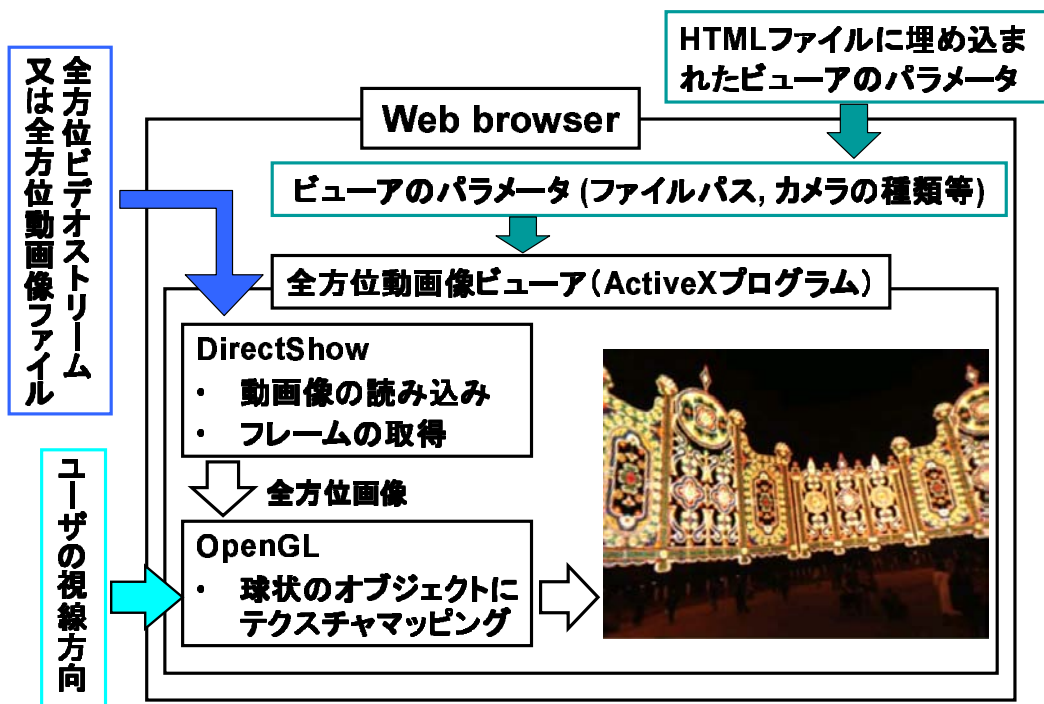


図 19 全方位動画ビューアの構造

2. 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動画のファイルパスの情報を基にサーバにアクセスし、全方位動画の受信を開始する。全方位動画がローカルな PC 内に格納されているのであればローカル PC 内から読み出しを始める。画像の入力に 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、入力される全方位動画から 1 フレームずつ全方位画像を取得する。
3. 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カードのテクスチャメモリへ転送する。そして、球状のオブジェクトにその全方位画像をテクスチャマッピングする。そのテクスチャマッピングされた球の内側から外向きに視線方向・視野を設定し、OpenGL でレンダリングすることで全方位画像中の特定の方向の画像を生成する。また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。

- 最後に，レンダリングされた画像を Web ページ上に提示することで利用者はその画像を見ることが可能となる．また，利用者が HMD を使用して全方位動画像を見るときには，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，全方位動画像がフルスクリーンで提示される．

2.5 実験

ネットワークテレプレゼンスシステムを実装し，蓄積型動画像とライブ型動画像を用いた実験を行った．蓄積型動画像を用いた実験では，全方位マルチカメラシステムを用いて撮影された高解像度の全方位動画像を鑑賞することを目的とした．一方，ライブ型動画像を用いた実験では，全方位カメラ等を搭載した車両からの全方位動画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効率的に複数人が同時に動画像を鑑賞することを目的とした．

2.5.1 高解像度の蓄積型動画像を用いた実験

この実験では全方位マルチカメラシステム Ladybug により得られた全方位動画像を LAN 内の全方位動画像サーバに格納しておき，観賞用 PC (表 1 参照) でその映像の鑑賞を行った．この実験では，観賞用 PC の性能の制約から全方位動画像は解像度を 1344×672 に設定した．また，動画像のフォーマットは，映像部には符号化が高速である理由から MPEG1 を，音声部には MPEG1 Layer2 を用いた．利用者は全方位動画像コンテンツを Web ブラウザや HMD により鑑賞した．

図 20 に観賞用 PC の Web ブラウザ画面を示す．映像と音声は同期して再生され，提示画像は 30fps で更新された．そして，マウスで画面上をドラッグすることにより自由に視線方向を変更でき，簡単に全方位動画像コンテンツを鑑賞することが可能であった．

より臨場感の高い鑑賞方法として，図 21 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

表 1 蓄積型動画鑑賞用 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

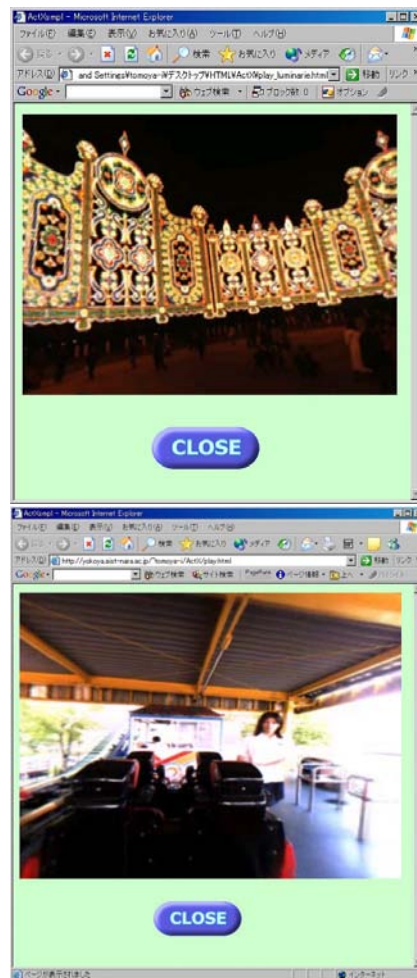


図 20 鑑賞用 PC の Web ブラウザ画面

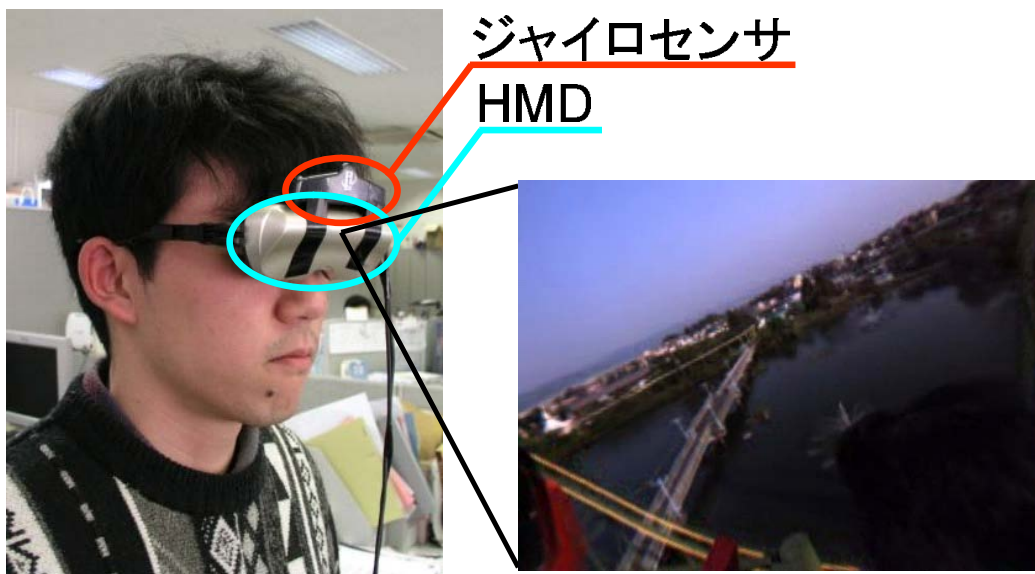


図 21 ジャイロセンサを取り付けた HMD

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。姿勢センサ付き HMD を装着し鑑賞を行ったところ、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を感じることなく 360 度視線方向を自由に変えることができた。

2.5.2 車載全方位カメラとマルチキャストを用いたライブ型動画像による実験

ライブ型動画像を用いた実験では、図 22 に示すように、全方位カメラ Hyper-Omni Vision を搭載した車両、全方位画像を取得・符号化・マルチキャスト配信する PC、全方位動画像ビューア、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 に画像を送る。画像取得圧縮符号化・マルチキャスト配信用 PC では、全方位動画像（解像度:640 × 480、フレームレート:24fps）を WindowsMediaEncoder[35] によって WindowsMedia 形式（ビットレート:832Kbps）に符号化する。符号化された動画像は、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 23,24 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動画像は鑑賞用 PC 上で動作する全方位動画像ビューアを用いて鑑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に対応していれば多数の利用者がビューアにより全方位動画像を受信することが可能である。

この実験では、4 台の PC で全方位動画像を鑑賞しようとしたところ、4 台ともそれぞれ任意の方向の画像を見ることができた。それらの PC の全方位動画像ビューアの画面例を図 25 に示す。画像は送信した動画像と同様の 24fps で提示され、カメラの撮影と画像提示の間には約 10 秒の時間遅延があった。時間遅延の主な原因は、全方位カメラからの画像を符号化するためのバッファリングによるものと、全方位動画像ビューアが符号化された全方位動画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、画像を受信するシステム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。

全方位動画像ビューア

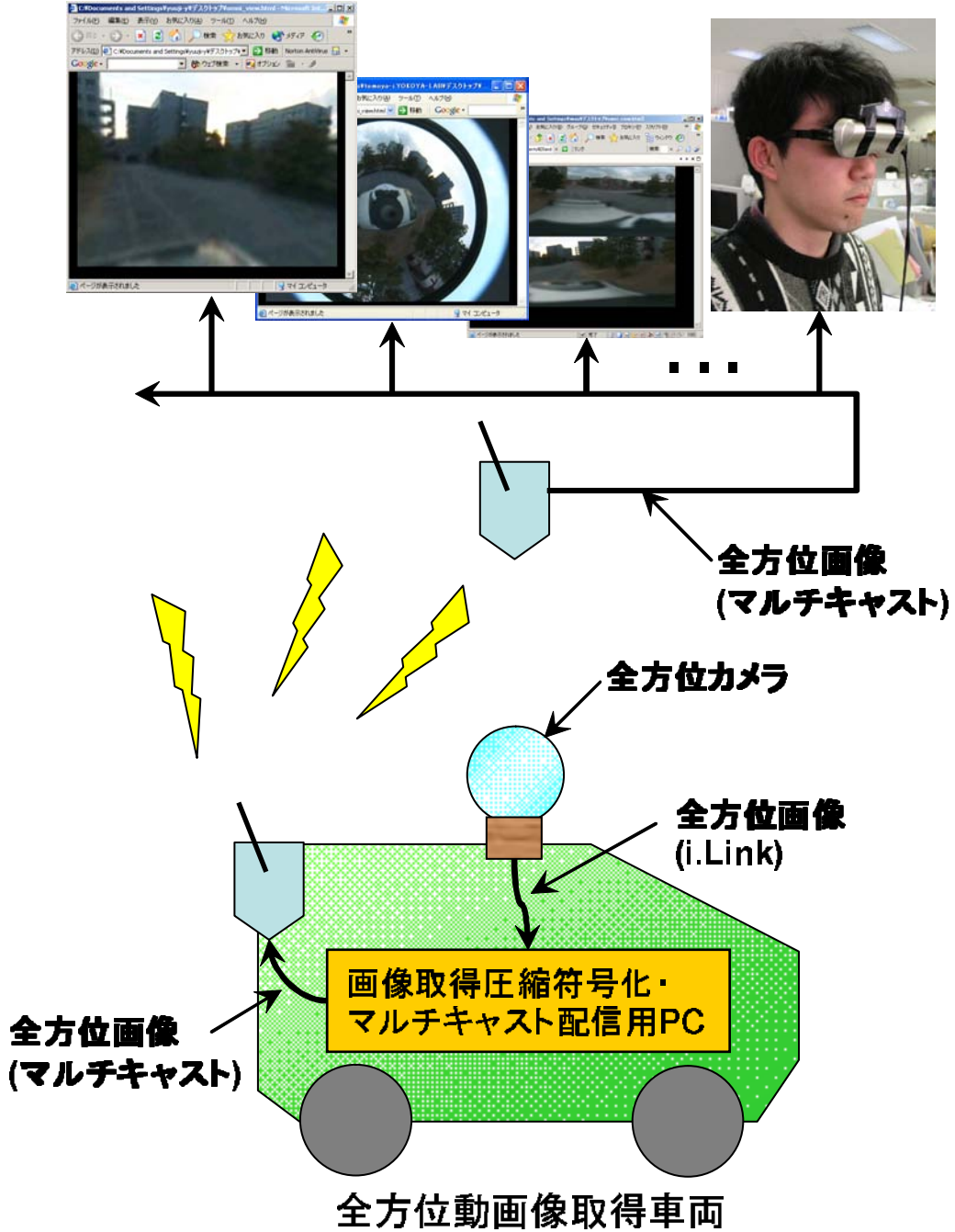


図 22 ライブ型動画像による実験のシステム構成

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



図 23 画像取得車両の外観



図 24 画像取得車両の内部

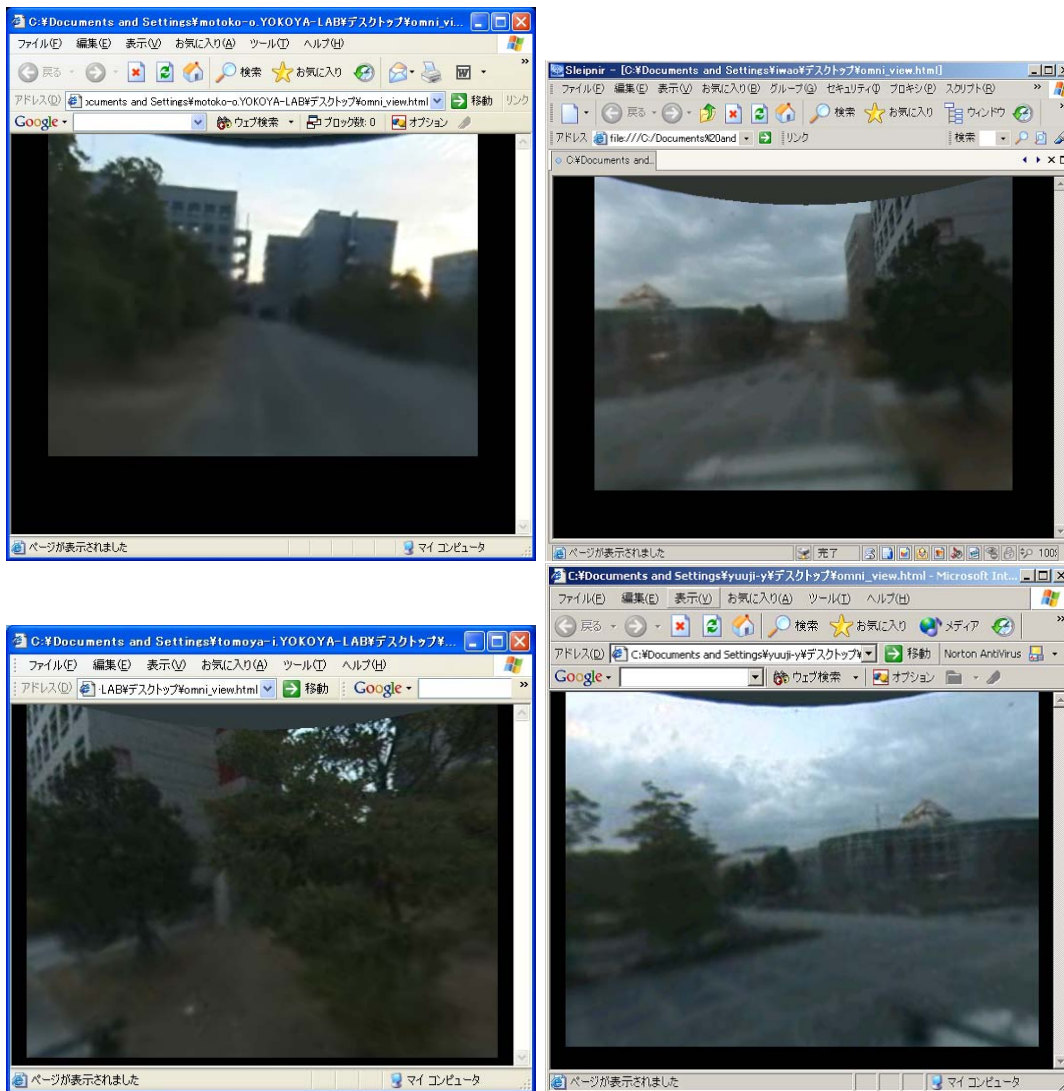


図 25 利用者それぞれの全方位動画像ビューアの画面例

2.6 考察

蓄積型全方位動画像を用いた実験では，利用者は高解像度な全方位動画像コンテンツを鑑賞することができた．鑑賞用 PC の性能による制約から全方位動画像の解像度を 1344×674 に落とす必要があった．Ladybug から池田らの手法により得られる全方位動画像の最大解像度は 3840×1920 であり，今後 PC の性能が向上することによりこのような高解像度の全方位動画像も扱えるようになると思われる．

ライブ型動画像を用いた実験では，全方位動画像の取得に用いた全方位カメラは NTSC の解像度で全方位の画像を取得するため，平面透視投影変換した画像に粗さが目立ち没入感を損なっていた．一方，蓄積型動画像の取得に用いた Ladybug は構造上，解像度の高い画像が得られるものの実時間で全方位動画像を配信するには適していなかった．そこで今後は，HD カメラを用いた HyperOmni Vision やライブ映像を取得可能なマルチカメラシステムを使用することにより高解像度で臨場感の高いネットワークテレプレゼンスを目指す．さらに，ライブ型動画像の実験では，カメラの撮影から画像提示の間に約 10 秒の時間遅延があり，遠隔地の相手とのコミュニケーション等の用途には利用が困難であると考えられる．今後は時間遅延の少ない映像符号化・伝送方法の利用も検討する．

3. 全方位画像を用いた動的環境における自由視点画像生成

本章では、多地点に配置された全方位カメラから得られるライブ動画を静的領域と動的領域に分離したハイブリッド手法による自由視点画像生成手法を提案する。そして、プロトタイプシステムを構築し、実際に自由視点テレプレゼンスが可能であるかを確認する。

3.1 静的領域と動的領域を分離したハイブリッド手法の概要

図 26 に静的領域と動的領域を分離したハイブリッドな自由視点画像生成手法の処理の流れを示す。本手法では環境に全方位カメラを複数台配置し、それらのカメラから得られる画像を用いて自由視点画像生成を行う。まず、それぞれのカメラからの画像に対し、背景差分処理により画像を静的領域と動的領域に分離する。静的領域に対しては Morphing、動的領域に対しては Visual Hull を用いてそれぞれ自由視点での画像を生成し、それらを合成した画像を最終的な自由視点画像として提示する。

以下、3.2 節では、提案手法で用いる全方位カメラについて述べ、3.3 節では、全方位動画を静的領域と動的領域に分離する処理について述べる。3.4 節と 3.5 節では、静的領域および動的領域に対しての自由視点画像生成について述べ、3.6 節で、静的領域と動的領域の自由視点画像の合成について述べる。そして、3.7 節と 3.8 節で提案手法を用いたプロトタイプシステムの実験とその考察を述べる。

3.2 全方位画像カメラによる画像取得

提案手法では、多地点に全方位カメラを固定して配置し、それらのカメラから得られる全方位動画を画像を用いて動的環境の自由視点画像を生成する。使用する全方位カメラには HyperOmni Vision を選んだ。これは、目標とするシステムがライブ動画を扱える必要があるため、Ladybug のようにオフライン処理を必要とする全方位カメラは適当でないと考えたためである。また、本手法ではカメラは

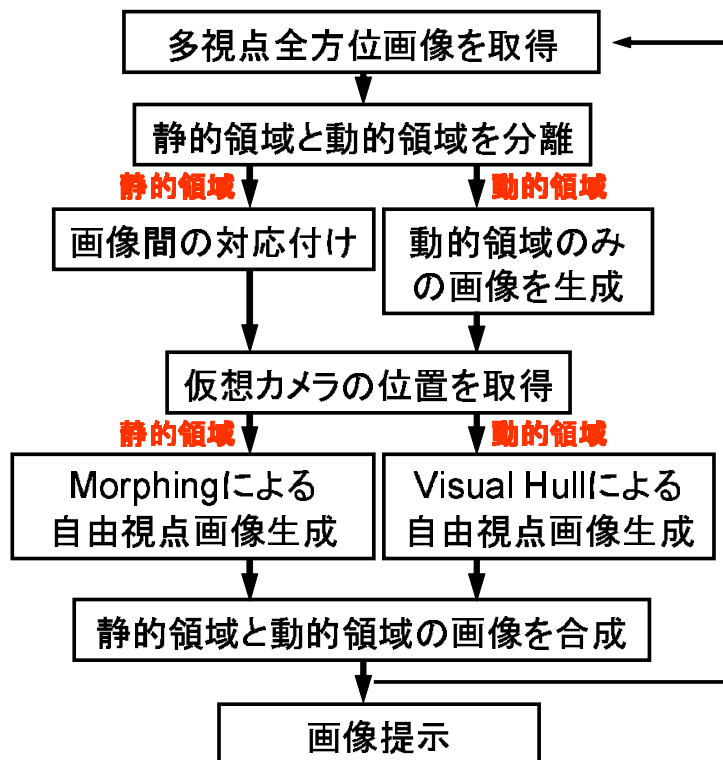


図 26 提案手法の処理の流れ

固定して設置されており，位置・姿勢はメジャーやレーザーレンジファインダで計測し既知であるものとする．

3.3 静的領域と動的領域の分離

複数の全方位カメラから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処理を行い，動的領域のみが含まれる全方位画像を生成する．この処理を行う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渡っての自由視点画像生成は困難である．そこで本研究では，蛍光灯や CRT ディスプレイのフリッカー，緩やかな照明変化，カメラに依存するノイズ等を考慮した

背景差分処理法 [33] を用いる．また，後の処理を高速に行うための情報として動物体の存在する方位角範囲を計算する．

3.3.1 背景更新を行う背景差分処理による動的領域の検出

本研究で用いる背景差分処理は，時系列に入力される画像の各画素について平均輝度値や分散値を保持し，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であれば背景とみなし，それ以外の場合は動的領域とみなす．さらに，動的領域と判断された領域に対して，その領域が新たに環境に加わった静的領域である場合を考慮して，ある程度時間が経つと背景と判断するよう背景画像の輝度値を変化させる．画像中の各背景画素の輝度 I は以下の式でモデル化する．

$$I = \bar{I} + \sigma \sin(2\pi\omega t) + k\xi \quad (3)$$

\bar{I} は輝度の時間平均， σ は輝度の振幅， ω は輝度の周波数， k は $-1 \sim +1$ までの値をとる係数， ξ 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 ディスプレイなどのフリッカー，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 は以下の範囲をとる得る．

$$\bar{I} - \sigma - \xi \leq I \leq \bar{I} + \sigma + \xi \quad (4)$$

入力画素の輝度が式 (4) の範囲内であればその画素は背景画素とみなし，そうでなければ動的領域とみなす．以下に処理手順を示す． ξ は予め動的領域の存在しない環境で測定しておく必要があるが，環境によって変更する必要はない．以下に時系列で入力される画像に対する処理手順を示す．

1. N フレームまでは背景画素の生成を行う． N フレームまでの画像を用いて各画素の輝度の平均値と分散を求め，背景画像とする． N は経験的に設定する値である．また， N フレーム目までは動的領域の検出処理は行わない．
2. $N + 1$ フレームからは動的領域の検出と背景画像の更新を行う．現フレームの各画素が式 4 の範囲内であれば背景画像の輝度の平均値と分散値を更

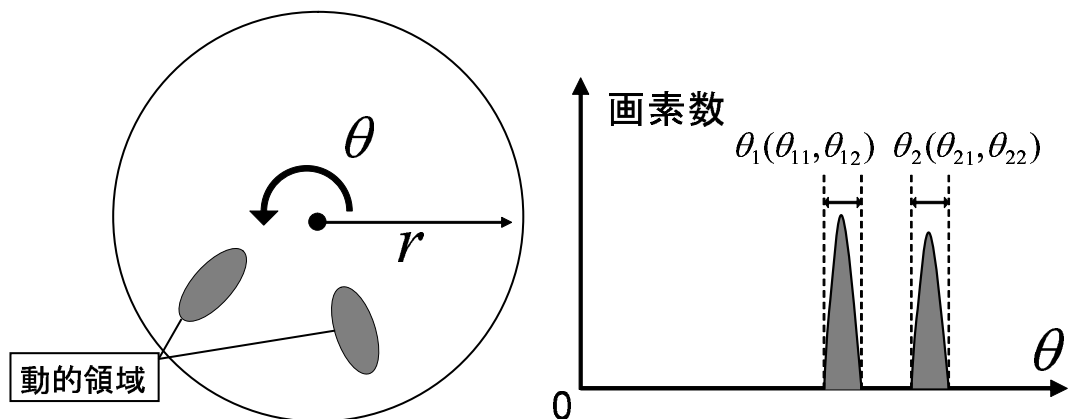


図 27 極座標変換によるヒストグラムの作成

新する．範囲外であればその画素は動的領域と判断し，現フレームの偏差に重みを加えて分散値のみを更新する．

以上の処理により，入力される画像から動的領域のみが含まれる画像を生成する．

3.3.2 動物体の方位角情報の取得

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, θ) に変換し， θ 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 27 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は，後の処理で使用するため保持しておく．

3.4 静的領域に対する Morphing を用いた自由視点画像生成

静的領域に対して Morphing を行い自由視点の全方位画像を生成する冨手らの手法 [30] を用いる．この手法は環境中の複数の地点で全方位画像を取得し，画像間の対応を与えることにより高速に自由視点における全方位画像を生成する．以下にその処理手順を示す．なお，全方位カメラの位置・姿勢および入力画像間の

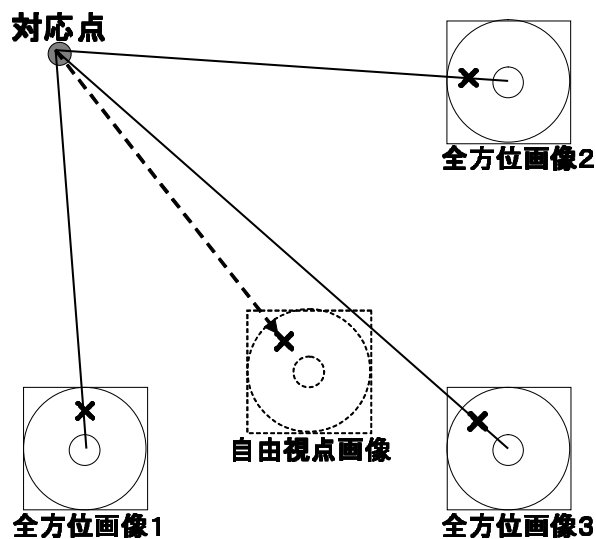


図 28 対応点の自由視点画像への投影

対応は既知とする。

1. 与えられた各対応点について，それらの点の三次元位置を全方位ステレオにより計算する。
2. 1. の処理で得られた三次元点を自由視点画像上に投影する（図 28 参照）。
3. 投影された点群に対してデローニー三角形分割を行い，三角形パッチを生成する。
4. 3. 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，自由視点位置に基づいて算出される重みを用いてブレンドし，自由視点画像における全方位画像とする。

3.5 動的領域に対する Visual Hull を用いた自由視点画像生成

本節では，3.3 節で述べた処理により生成された動的領域のみが含まれる画像を用いて動物体の自由視点画像生成を行う。また，その処理の計算量削減方法について述べる。

3.5.1 Visual Hull の算出処理

動的領域に対する自由視点画像生成には Visual Hull を用いる。Visual Hull とは、多視点から撮影された物体のシルエットとカメラの光学中心によって形成される視体積の積集合空間を指す。この Visual Hull を表現する手法として視体積交差法 [14] がある。しかし、視体積交差法では空間をボクセルで表現するため、対象とする空間が広がるとそのデータ量が膨大となる。そこで、本研究では Visual Hull を表現する手法に Image-Based Visual Hull [24] 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対して、仮想カメラの光学中心（自由視点）とその画素を結ぶ光線が 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。その手順を以下に示す。

1. 仮想カメラの画像面のある画素について、仮想カメラの光学中心とその画素を結ぶ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仮想カメラの光学中心がエピポール、直線がエピポーラ線となる。
2. 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領域を調べ、全てのエピポーラ線が動的領域と交差する領域（共通領域）があるならば、仮想カメラからの直線は Visual Hull を貫通する。共通領域が無い場合は、Visual Hull を貫通しないと判断できる（図 29 参照）。
3. 仮想カメラからの直線が 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、共通領域の始点への各実カメラからのベクトルと自由視点からの直線ベクトルとを比較し、最もベクトルの向きが近い実カメラの画素の色を対象としている画素の色とする。
4. 1. ~ 3. の処理を自由視点画像上の全ての画素に対して行う。

3.5.2 動物体の位置推定による処理の削減

前節では、Visual Hull の計算処理について述べた。この処理を自由視点画像上の全ての画素について行うと計算量が多く、実時間での自由視点画像生成は困難

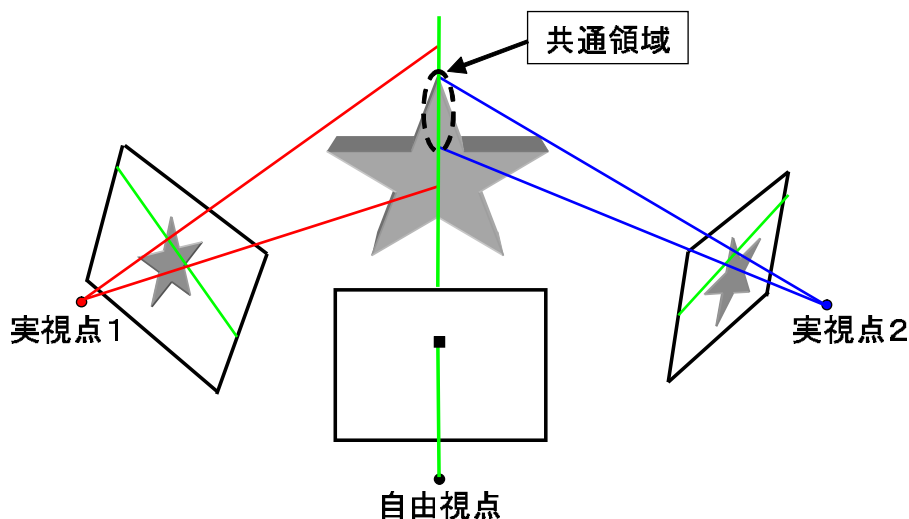


図 29 共通領域の有無の判定

となる．そこで，計算量を削減するために，予め動物体の環境中での位置を推定し，その位置を仮想カメラの画像面上に投影し，計算が必要な動的領域を限定する．動物体の位置の計算には寺沢らの手法 [36] を用いる．以下では，説明を分かりやすくするためにカメラを二つ使った場合の動物体の位置推定について述べる．

まず，3.3.2 項で得られた方位角範囲から動物体の存在領域を推定する．図 30 より，動物体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．

(交点座標) = (カメラの座標値) + (単位ベクトル) × (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．

$$\begin{aligned}
 (\text{交点の } x \text{ 座標}) &= S_{1x} + V_{1x} \times M_1 \\
 &= S_{2x} + V_{2x} \times M_2 \\
 (\text{交点の } y \text{ 座標}) &= S_{1y} + V_{1y} \times M_1
 \end{aligned} \tag{5}$$

$$= S_{2y} + V_{2y} \times M_2 \quad (6)$$

式 (5),(6) より，各媒介変数は以下の式で求めることができる（図 31 参照）．

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

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

各カメラから伸ばした全ての 2 直線について以上の処理を行い，動物体の存在領域を推定する．動物体の存在領域が分かればその領域を仮想カメラの画像面上に投影し，投影された範囲の画素に対してのみ Visual Hull の処理を行う．

3.6 静的領域と動的領域の自由視点画像の合成

3.4, 3.5 節の処理により，ある自由視点位置における静的領域の全方位画像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．

静的領域と動的領域の自由視点画像は別々の手法により生成されるが，静的領域に対して用いる自由視点画像生成手法は，環境の三次元形状を疎に持っているものと考えられる．また，動的領域に対して用いる自由視点画像生成手法は自由視点位置から物体までの距離を推定していることと等価である．よって，3DCG オブジェクトを配置するかのように二つの全方位画像を合成しても環境中の物体の位置関係に違和感を感じることはない．

3.7 実験

静的領域と動的領域を分離したハイブリッド手法により自由視点テレプレゼンスを可能にするプロトタイプシステムを実装し，ライブ動画像を入力とした実時間自由視点画像生成を行った．図 32 に示すような室内環境に，全方位カメラを 3 台設置し，それらを接続した PC で自由視点画像生成を行った．

プロトタイプシステムによる実験では，実験を簡略化するために Morphing に用いる画像は予め静止環境を撮影したものを使用し，入力される映像を使用して

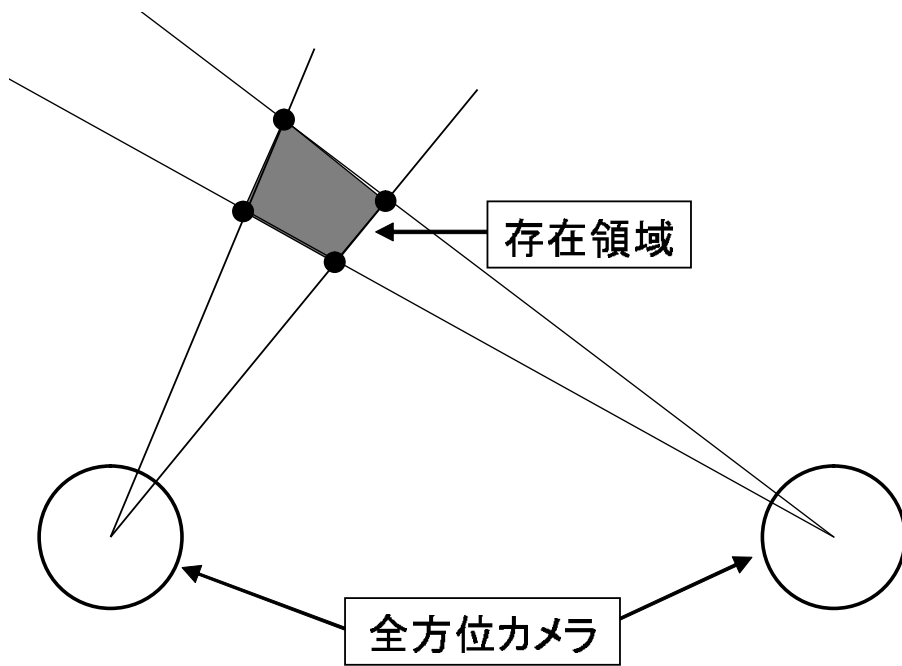


図 30 推定される存在領域

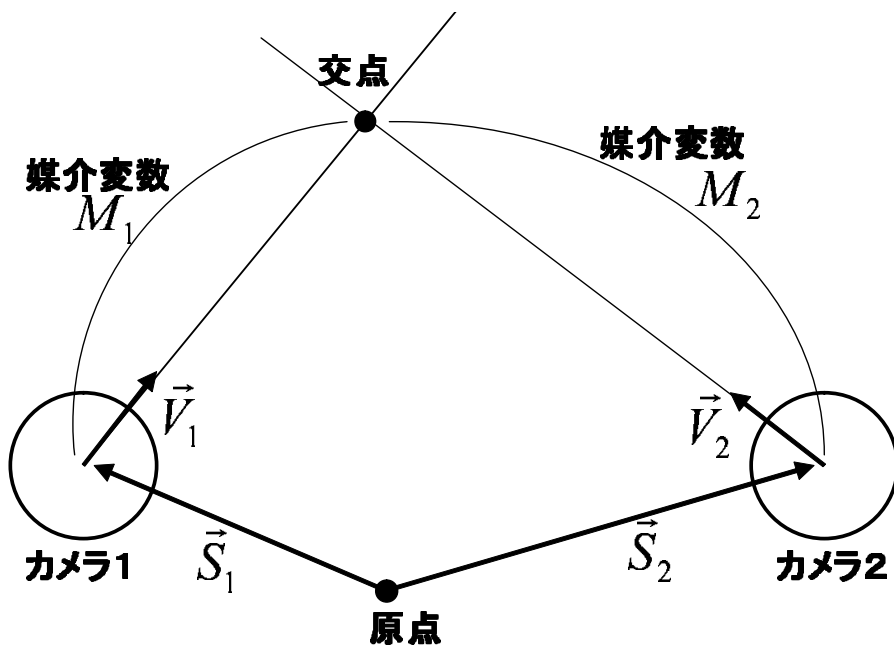


図 31 直線の媒介変数表示

画像を更新する処理は行わないこととした。そして、静的環境の対応点は予め手動で与えるものとし、全方位カメラの位置・姿勢は既知とした。また、Morphing による制限から自由視点画像を生成できる範囲は各実カメラを結ぶ平面上に限定した。3.5 節において Image-Based Visual Hull の計算量削減について述べたが、その計算量削減処理に加えて、利用者が見ている範囲の画像のみを処理することとした。全方位カメラを使用する利点として、同じ全方位画像を用いて複数の利用者がそれぞれ別の方向を見ることができるとのこと上げられるが、自由視点でのテレプレゼンスにおいては、それぞれの利用者がそれぞれ違う視点位置から見ることが想定されるため、利用者が見ている範囲のみの自由視点画像生成をすることは妥当だと考える。また、動的領域の自由視点画像については穴の除去のためにメディアンフィルタを適用した。

3.7.1 システム構成

図 33 にプロトタイプシステムの構成を示す。IEEE1394 で接続可能な全方位カメラを 3 台設置し、それらをすべて 1 台の画像取得・生成用 PC に接続した。使用した全方位カメラは未陰産業製 SOIOS 55-Cam である。この全方位カメラは、640 × 480 の解像度の動画像を最高 15fps で取得可能であり、垂直方向には仰角 12 度、俯角 50 度、水平方向には 360 度の視野を持つ。カメラ間の距離は 1m とした。画像取得・生成用 PC は CPU に Pentium4-3.2GHz、メモリを 2GB、グラフィックスカードは RADEON9800XT を搭載している。システムの利用者は画像取得・生成用 PC のディスプレイで生成画像を確認する。また、利用者はマウスのドラッグ操作により視線方向を変更し、キーボード操作で視点位置を変更することができる。

3.7.2 結果

図 34 に実験環境の様子と自由視点位置における生成画像を示す。図 34 は左から自由視点位置と視線方向、実験環境の様子（男女が歩行）、自由視点全方位画像、自由視点全方位画像を平面透視投影変換した画像である。多地点全方位画

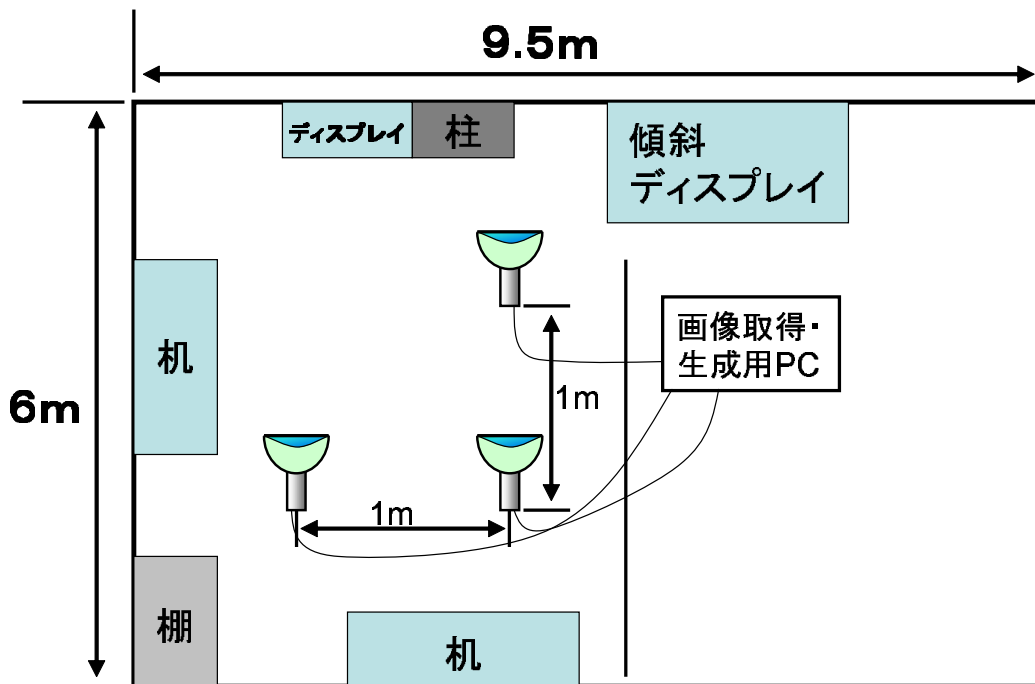


図 32 自由視点画像生成の実験環境

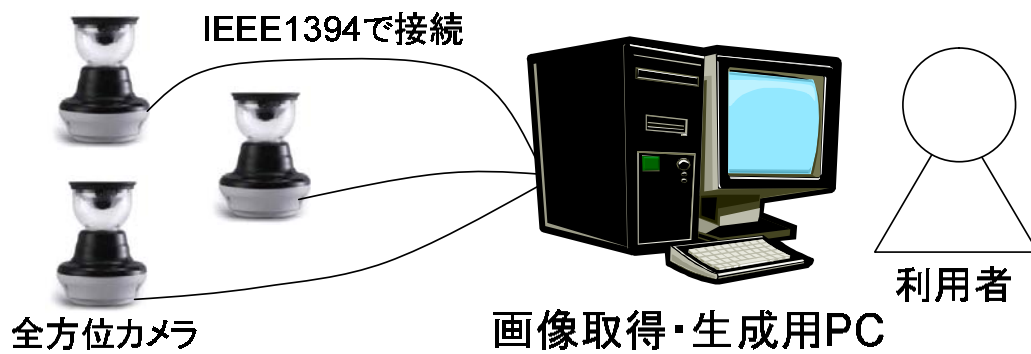


図 33 プロトタイプシステムの構成

像の入力から平面透視投影変換された画像の提示まで約 200ms (約 5fps) で処理することができた。内訳は、同期した画像の取り込みと背景差分処理に約 50ms、Morphing に約 50ms、Image-Based Visual Hull とメディアンフィルタに約 100ms となっていた。Image-Based Visual Hull の処理については、自由視点画像上に投影される動的領域の大きさに比例して計算時間が変化する。そのためフレームレートは一定ではなく、場合によっては画像提示が 3fps に低下した。また、生成した自由視点画像と現実環境を比較すると、動物体の位置が 50cm 程度ずれていた。

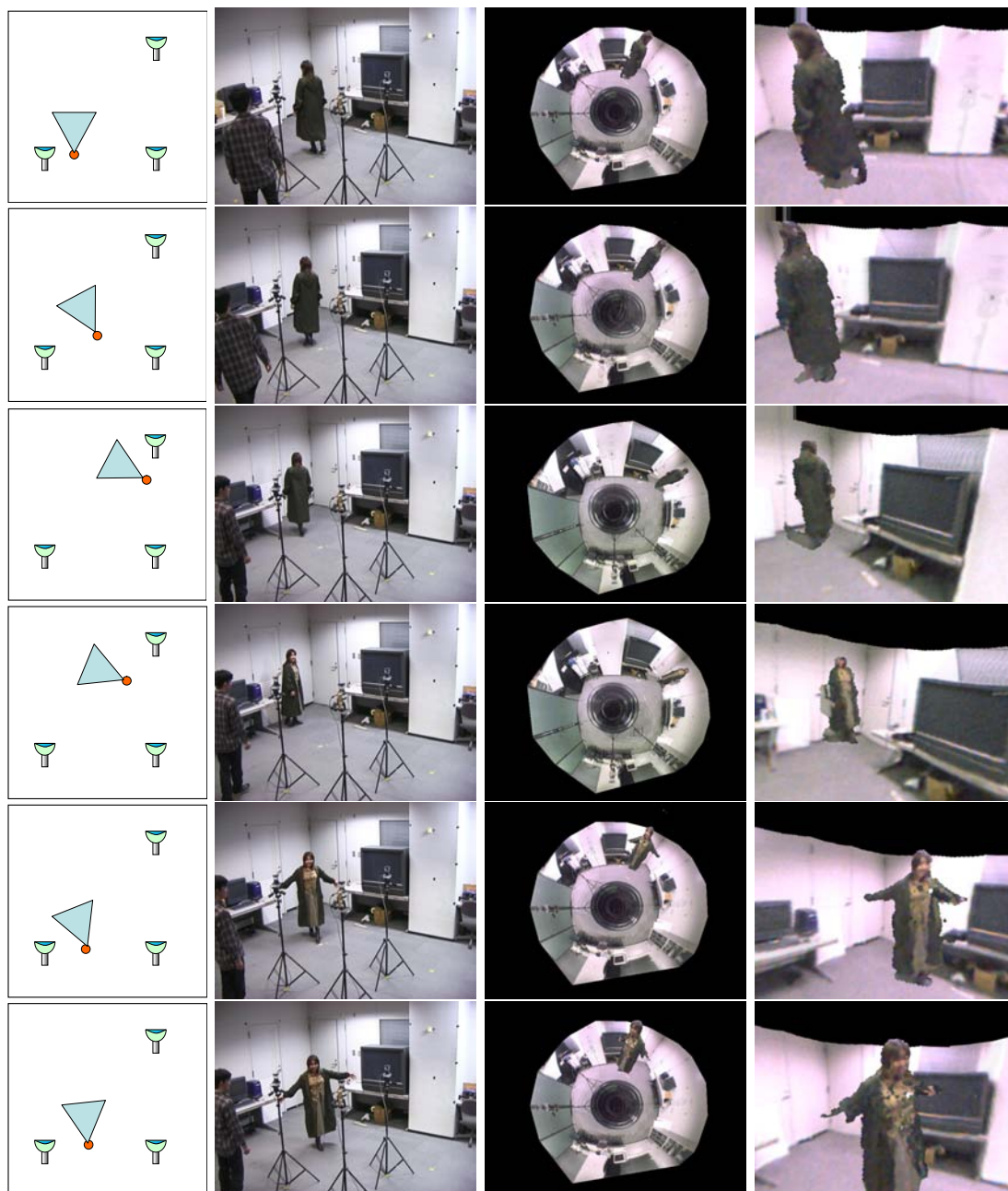


図 34 実験環境の様子と生成された自由視点画像

3.8 考察

図 34 の生成した自由視点全方位画像では男性は映っていない。これは動的領域は利用者が見ている範囲のみ計算するためである。冨手らの手法により静的領域の全方位画像を一度に生成できるが、利用者が見ている方向以外の画像を生成するのは無駄な処理となるので、より高速に自由視点画像生成を行うためには無駄な計算を省く方法を検討する必要がある。自由視点画像上で静的領域と動的領域の位置ずれが起こっているが、これは実カメラを設置した際の位置・姿勢のずれやオクルージョンの影響などにより静的環境の対応点数が少ないために Morphing によって生成した画像に歪みを生じていることが原因として考えられる。この問題の解決にはカメラの位置・姿勢を精度よく推定すること、マルチベースラインステレオ [11] 等を利用し対応点数を増やすことが考えられる。また、今回の実験のカメラ配置では相互のカメラによりオクルージョンがあり、動的領域を正しく検出できない場合があった。環境に対するカメラの配置方法を検討する必要がある。

4. まとめと今後の課題

本研究では、「広範囲な現実環境の撮影」、「インタラクティブな画像提示」、「ライブ動画像の処理と提示」、「複数の利用者に対する画像提示」の機能を有するテレプレゼンスシステムの実現を目標とし、そのためにまず二つのシステムを構築した。一つは、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利用できる固定視点テレプレゼンスシステム、もう一つは静的領域と動的領域を分離したハイブリッド手法による自由視点テレプレゼンスシステムである。これらのシステムを実装することにより、目標とする機能を有するテレプレゼンスシステムが実現可能であるかの確認を行った。

まず、固定視点テレプレゼンスでは、システムのネットワーク化により全方位動画像を用いたテレプレゼンスを複数人が同時に利用できるネットワークテレプレゼンスシステムを実現した。そして、本システムを用いることにより、蓄積型動画像の実験では高品質な動画像を利用できることを確認し、ライブ型動画像の実験では無線・有線ネットワーク上でマルチキャスト配信を行い、複数の利用者が同時にそれぞれ任意の方向の動画像を鑑賞するインタラクティブTV放送のような利用が可能であることを確認した。さらに、利用者が容易にコンテンツを鑑賞・操作できることを確認した。

次に、全方位動画像を静的領域と動的領域に分離し、静的領域には Morphing、動的領域には Visual Hull を適用することで、広範囲の動的環境を自由な視点位置で見回すことができる自由視点画像生成手法を提案した。さらに、提案手法を用いたプロトタイプシステムを実装し、自由視点画像生成が可能であることを確認した。そして、自由視点画像生成の計算処理を削減することによりライブ動画像を用いた実時間自由視点画像生成が可能であることを実験により示した。この手法の特徴として、モデルベースドレンダリングとイメージベースドレンダリングの特徴を併せ持つハイブリッド手法を利用しており、幾何学的情報を持つことができるため仮想物体を仮想化された実環境中に配置することも可能である。

本研究では、ライブ動画像を扱うために HyperOmni Vision を用いた。HyperOmni Vision は一般的なカメラに曲面ミラーを取り付けることで全方位の撮影を可能にしている。そのため、全方位画像の一部を切り出し平面透視投影変換する

と解像度が低くなり、利用者が画像を見た際に臨場感を損なっていた。より臨場感の高いテレプレゼンスを行うためには HyperOmni Vision に HD カメラを使用することやライブ動画像を取得可能な全方位マルチカメラシステムを導入することが考えられる。

ネットワークテレプレゼンスの実験において、全方位動画像を符号化しマルチキャスト配信した場合、全方位カメラによる画像の取得から全方位動画像ビューアでその画像が提示されるまでの間に約 10 秒の時間遅延があった。そのため、遠隔地の相手とのコミュニケーション等の用途には利用が困難であると考えられる。伝送遅延の少ない映像伝送方法として DVTS があり、今後は DVTS をシステムに組み込むことにより、この問題を解決したいと考えている。

また、提案した自由視点画像生成手法について、プロトタイプシステムによって生成した自由視点画像は環境の幾何学的情報を持つので物体の位置関係を把握することができたが、両眼立体視可能なほどの精度はなかった。今後は精度の向上により両眼立体視可能な自由視点画像生成を目指す。

今後の大きな課題は、提案したネットワークテレプレゼンスシステムと自由視点テレプレゼンスシステムを融合させることで目標とするテレプレゼンスシステムを実現することである。そのためには、自由視点テレプレゼンスにおいて、環境中に配置した全方位カメラをネットワークで接続する必要がある。そして、自由視点画像生成に使用する全方位カメラを選択し、それらの全方位カメラから同期した画像を取得しなければならない。このテレプレゼンスシステムを実現するために、ネットワーク上で PC 間の時刻同期を行うプロトコルを使用して全方位カメラとそれを接続した PC の時刻同期を行い、全方位画像を取得した時間のタイムコード付き全方位画像を同期画像生成サーバに集める。そして、自由視点画像生成を行う PC がサーバにどの全方位カメラからの画像を必要とするかのリクエストを送信し、同期した全方位画像を得て自由視点画像生成を行う予定である。また、このサーバから全方位画像を送信する際には、マルチキャストプロトコルを使用することにより効率的に全方位画像を送信できると考えている。

謝辞

本研究を進めるにあたり，適切なご指導を頂いた視覚情報メディア講座 横矢直和教授に厚くお礼申し上げます．副指導教官としてご助言を頂いた情報科学センター 砂原秀樹教授ならびに視覚情報メディア講座 山澤一誠助教授に深く感謝致します．多くの御助言を頂いた視覚情報メディア講座 神原誠之助手，佐藤智和助手に心からお礼申し上げます．また，研究活動を行うにあたり，ご指導，ご協力頂いた視覚情報メディア講座の学生の皆様，ならびに守屋知代女史に厚く感謝致します．

参考文献

- [1] Ed. S. Moezzi. Special issue on immersive telepresence. *IEEE MultiMedia*, Vol. 4, No. 1, pp. 17–56, 1997.
- [2] Z. L. Cao, S. J. Oh, and E. L. Hall. Dynamic omnidirectional vision for mobile robots. In *Proc. SPIE Intelligent Robots and Computer Vision*, 第 579 卷, pp. 405–414, 1985.
- [3] Y. Yagi, Y. Nishizawa, and M. Yachida. Estimating location and avoiding collision agaist unknown obstacle for the mobile robot using omnidirectional image sensor copic. In *Proc. Int. Workshop on Intelligent Robots and System*, pp. 909–914, 1991.
- [4] 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.
- [5] 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. 14th IAPR Int. Conf. on Pattern Recognition (ICPR'98)*, pp. 485–489, 1998.
- [6] 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.
- [7] 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.
- [8] 小野晋太郎, 川崎洋, 池内克史. 時空間距離画像の提案・考察と街のモデル化への応用. 情報処理学会 コンピュータビジョンとイメージメディア研究会 CVIM142-010, 2004.

- [9] 浅井俊弘, 神原誠之, 横矢直和. 全方位レンジデータと全方位画像の統合による屋外環境の三次元モデル化. 画像の認識・理解シンポジウム (MIRU2004) 講演論文集, pp. 195–200, 2004.
- [10] 佐藤智和, 池田聖, 横矢直和. 複数の動画像系列を用いたマルチベースラインステレオによる屋外環境の三次元復元. 情報科学技術フォーラム (FIT) 情報技術レターズ, Vol. 1, pp. 151–152, 2002.
- [11] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 353–363, 1993.
- [12] 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.
- [13] P. Rander, P. J. Narayanan, and T. Kanade. Virtualized reality: Constructing time-varying virtual worlds from real world events. In *Proc. of IEEE Visualization'97*, pp. 277–283, 1997.
- [14] T. Matsuyama and T. Takai. Generation, visualization, and editing of 3d video. In *Proc. 1st Int. Symposium on 3D Data Processing, Visualization and Transmission*, pp. 234–245, 2002.
- [15] 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.
- [16] S. Yamazaki, K. Ikeuchi, and Y. Shingawa. Determining plausible mapping between images without a priori knowledge. In *Proc. Asian Conference on Computer Vision 2004 (ACCV2004)*, pp. 408–413, 2004.
- [17] S. E. Chen and L. Williams. View interpolation for image synthesis. In *Proc. of SIGGRAPH'96*, pp. 279–288, 1996.

- [18] S. M. Seitz and C. R. Dyer. View morphing. In *Proc. of SIGGRAPH'96*, pp. 21–30, 1996.
- [19] 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.
- [20] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. of SIGGRAPH'96*, pp. 31–42, 1996.
- [21] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. of SIGGRAPH'96*, pp. 43–54, 1996.
- [22] 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.
- [23] 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.
- [24] W. Matusik, C. Buehler, S. Gortler, and L. McMillan. Image-based visual hulls. In *Proc. of SIGGRAPH2000*, pp. 369–374, 2000.
- [25] 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.
- [26] M. Irani, T. Hassner, and P. Anandan. What does the scene look like from a scene point? In *Proc. 7th European Conf. on Computer Vision*, pp. 1176–1183, 2003.
- [27] T. Koyama, I. Kitahara, and Y. Ohta. Live mixed-reality 3d video in soccer stadium. In *Proc. 2nd IEEE/ACM Int. Symposium on Mixed and Augmented Reality (ISMAR03)*, pp. 987–990, 2003.

- [28] 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.
- [29] 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. 15th Int. Conf. on Pattern Recognition*, pp. 463–467, 2000.
- [30] K. Tomite, K. Yamazawa, and N. Yokoya. Arbitrary viewpoint rendering from multiple omnidirectional images. In *Proc. 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002)*, pp. 987–990, 2002.
- [31] S.E.Chen. Quicktime vr - an image-based approach to virtual environment navigation. In *Proc. SIGGRAPH'95*, pp. 29–38, 1995.
- [32] 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*, Vol. 71, No. 2, pp. 154–165, 1998.
- [33] 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.
- [34] T. Yamamoto and M. Doi. Panovi: Panoramic movie system for real-time network transmission. In *IEEE 4th Workshop on Multimedia Signal Processing*, pp. 389–394, 2001.
- [35] Microsoft Corporation. Windows media encoder 9 series. <http://www.microsoft.com/windows/windowsmedia/9series/encoder/default.aspx>.
- [36] 寺沢征彦, 山澤一誠, 竹村治雄, 横矢直和. 複数の全方位画像センサを用いた遠隔監視システムにおける複数移動物体の存在領域推定. 電子情報通信学会

技術研究報告 PRMU2000-195, 2001.