

複数の全方位動画像を用いた自由視点テレプレゼンス

Telepresence Using Multiple Omni-directional Videos

学生会員 石川 智也[†], 山澤 一 誠[†], 正会員 横 矢 直 和[†]

Tomoya Ishikawa[†], Kazumasa Yamazawa[†] and Naokazu Yokoya[†]

Abstract The advent of high-speed network and high performance PCs has prompted research into networked telepresence, which allows a user to see a virtualized real scene in remote places. View-dependent representation, which provides a user with arbitrary images using an HMD or an immersive display, is especially effective in creating a rich telepresence. The goal of this study is to create a novel view telepresence that enables a user to control the viewpoint and view-direction by virtualizing real dynamic environments. We describe a novel method of generating views that uses image-based rendering techniques from multiple omni-directional images captured from different positions and that evaluates image quality using a simulated environment. We also describe our prototype system and an experiment with the novel view telepresence that used the system in a real environment. Our prototype novel view telepresence system constructs a virtualized environment from real live videos. The system synthesizes a view based on the user's viewpoint and view-direction as measured by a magnetic sensor attached to an HMD and presents the generated view on the HMD. Our system can generate a user's view in real-time by presenting corresponding points and estimating camera parameters in advance.

キーワード：自由視点 テレプレゼンス 全方位画像 イメージベースレンダリング ネットワーク ライブ映像

1. ま え が き

近年のネットワークの高速化やPCの高性能化に伴って、利用者に対し時空間的に隔たった場所に実際に居るかのような、没入感覚を与え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。テレプレゼンスでは、遠隔地の現実環境を撮影した画像から仮想環境を構築するため、3DCGオブジェクトを用いて構築した環境に比べ、利用者を与える没入感が高いといった特徴がある。そのため、仮想旅行やゲーム等のエンターテインメント分野のみならず、教育・医療分野でのデジタルコンテンツとしても利用可能であり、その応用範囲は広い。

これまで我々は、テレプレゼンスの臨場感を高めるために、利用者が提示されている映像の視線方向を自由に変更可能な全方位動画像を用いたシステムを提案してきた¹⁾⁻³⁾。これらのシステムでは、周囲360度を一度に撮影可能な全方位カメラや全方位型マルチカメラシステムを用いて環境を撮影し、利用者の視線方向に応じた画像を提示することで、高臨場感のテレプレゼンスを実現している。さらに、イ

ンターネットを介した全方位映像の伝送により、遠隔地のライブ映像を自由な視線方向で観賞することも可能である³⁾。これまでのテレビ放送のように、カメラマンが撮影した方向の映像のみを見るのではなく、利用者が自由に見る方向を変えることができることから、これらのシステムは通信と放送が融合する時代におけるネットワークメディアとして有望視されている。本研究では、より高臨場感のテレプレゼンスの実現のために、動的実環境およびライブ映像に対応した視線方向と視点位置を自由に変更可能なテレプレゼンスシステムの実現を目的とする。特に、利用者に没入感覚を与えるために、環境中に配置されたカメラ群の内側の仮想化のみならず外側の仮想化にも同時に対応し、撮影環境全体の仮想化を実現する。このシステムにより、遠隔の観光地に居るような感覚でその空間を自由に歩行するような仮想観光アプリケーションや、さらにその遠隔地に存在する人と対話可能なアプリケーションへの応用が可能となる。また本研究では、このようなアプリケーションへの応用を想定し、以下に従来研究について述べる。

1.1 関連研究

テレプレゼンスにおいて、最も有名なシステムの一つにChen⁴⁾によって提案されたQuickTimeVRがある。このシステムは、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パノラマ画像を生成し、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

2006年2月28日受付, 2006年4月25日再受付, 2006年6月5日採録

[†] 奈良先端科学技術大学院大学 情報科学研究科

(〒630-0192 奈良県生駒市高山町 8916-5, 0743-72-5296)

[†] Graduate School of Information Science, Nara Institute of Science and Technology

(8916-5, Takayama-cho, Ikoma, Nara, 630-0192 Japan)

グすることで、利用者に写実的な仮想環境の提示を行っている。このように環境を明示的に3次元復元することなく、撮影画像の変形・合成により仮想環境を構築・提示する技術はImage Based Rendering(IBR)と呼ばれ、計算コストの低さや写実性の高さから現在のテレプレゼンスの研究で広く利用されている。

一方、多地点で撮影された画像から環境を3次元復元することで仮想環境を構築・提示する技術は、Model Based Rendering(MBR)と呼ばれる。3-D Roomと呼ばれる多数のカメラが配置された室内環境で、動的イベントを仮想化するKanadeら⁵⁾や斎藤ら⁶⁾の研究はその代表例である。これらMBRを用いた手法では、複数のカメラ映像から3次元復元を行うため、非常に多くの計算コストがかかる。そのため、ライブ映像の提示は困難となる。また、3次元復元の精度が提示画像の品質に大きく影響することから、なるべく多くのカメラが同じ領域を撮影するようにカメラを配置するため、実際に提示可能な仮想環境は配置されたカメラの内側のごく僅かな範囲に限られる。テレプレゼンスにおいて、利用者の視線方向・視点位置以外の仮想環境を提示する必要はないことや、撮影環境全体の仮想化を考慮するとMBRよりもIBRが適している。したがって、以下ではIBRによる関連研究に焦点をあてて従来研究を概観する。

Uyttendaeleら⁷⁾は、全方位型マルチカメラシステムを用いて、利用者が視線方向や視点位置を変更可能なシステムを提案した。このシステムでは、利用者がジョイパッドによる操作でインタラクティブに視線方向や視点位置が変更可能である。しかし、変更可能な視点位置は撮影経路上に制限されている。

視点位置を自由に変更するためのIBR手法も様々提案されている^{8)~11)}。Koyamaら⁸⁾や稲本ら⁹⁾はサッカーシーンの特徴を利用することで、自由な視点位置での画像提示を可能にしている。Koyamaらの手法⁸⁾は、利用者の視点位置とサッカープレイヤは離れているという仮定に基づき、平面にプレイヤの画像をマッピングし、近似的に自由な視点での画像を提示する。稲本らの手法⁹⁾は、サッカーフィールドやゴールポストは複数の平面で構成されていることや遠景は平面と見做せることを利用し、サッカーフィールドとゴールポストに対してはHomography、遠景に対してはモザイクングと領域切り出しを行い、それらの画像を合成することで自由視点画像を生成する。これらの手法は、撮影環境が複数の平面で構成されていることや平面で近似できる必要があるため、サッカーシーンやそれに類似したシーン以外への適用は困難である。AliagaとCarlbom¹⁰⁾は、全方位カメラにより撮影環境中の光線情報をデータベース化し、利用者の視線方向・視点位置における画像面を通過する光線を、データベースからリサンプリングすることで自由視点画像を生成した。この手法では、静止環境を密に撮影する必要があり動的環境への対応は困難である。Zitnick

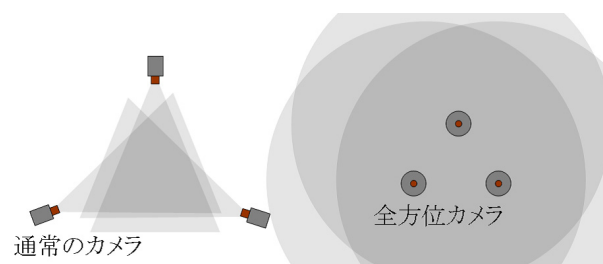


図1 通常のカメラ(左)と全方位カメラ(右)を用いて仮想化される領域

Areas virtualized with standard cameras (left) and omnidirectional cameras (right).

らの手法¹¹⁾は、各カメラについて距離画像を生成し、距離が類似した物体をレイヤー化し、利用者の自由視点位置における画像面に各レイヤーを投影することで画像を生成する。この手法では、距離画像生成の際にオフライン処理が必要になるので、ライブ映像への適用は困難である。

1.2 本研究のアプローチ

本論文では、利用者が遠隔地の動的環境中をライブかつ自由な視線方向や視点位置で観賞可能な高臨場感テレプレゼンスを提案する。そのためには、以下の機能が要求される。

- (A) 動的な撮影環境全体に対応
- (B) 遠隔地映像の伝送
- (C) ライブ映像からの自由視点画像生成
- (D) 利用者の動作に応じた画像提示

本研究では、(A) 通常のカメラを用いて仮想化されるカメラ群の内側の領域(図1左)のみならず、カメラ群の外側の動的領域にも対応するために、多地点に配置した全方位カメラにより環境を撮影する(図1右)。さらに、取得した全方位画像を静的領域と動的領域に分離し、ネットワークを介することで(B) 遠隔地へ各領域の画像を伝送する。伝送されたそれぞれの領域に対して、IBRに基づいた自由視点画像生成手法を適用することで(C) ライブ映像から高速に自由視点画像を生成する。また、利用者は位置・姿勢センサ付きHMDを装着することで、(D) 利用者の動作に応じた画像提示を可能とする。

要求機能(A)~(D)を満たすテレプレゼンスのための自由視点画像生成手法を提案し、その生成画像の品質をシミュレーション環境を用いて定量的に評価する。さらに、実験では実際にプロトタイプシステムを構築し、動的環境を撮影したライブ映像からの利用者の動作に応じた自由視点テレプレゼンスが可能であることを確認する。

以下、2節では本研究で用いる全方位カメラHyperOmniVisionの光学系とそのキャリブレーションについて述べる。3節では、多地点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。4節では、シミュレーション画像を用いて自由視点画像を生成し、生成画像の定量的な評価を行う。5節では、構築したプロトタイプシステムについて述べ、実験として動的環境を撮影したライブ映像をネットワークを介して伝送し、その映像から自由視点テレプレゼ

ンスを行う．最後に6節で，まとめと今後の課題について述べる．

2. 全方位カメラ

本研究で用いる全方位カメラ HyperOmniVision¹²⁾ には以下に示す特徴がある．

- ・ 光学特性が透視投影：全方位画像から平面透視投影画像に変換が可能．
- ・ 一度に全方位の画像を取得可能：継ぎ目のない全方位画像を使用することにより仮想環境を利用者に提示した際の違和感を低減可能．
- ・ ビデオレートでの撮影が可能：カメラ部分には一般的なビデオカメラが使用できるため，映像のビデオレート取り込みが可能．

これらの特徴は，目標とするテレプレゼンスシステムに適したものであるので，本研究では HyperOmniVision を採用する．また，この全方位カメラを多地点に配置することで，環境中に配置されたカメラ群の内側と外側の領域を同時に仮想化することが可能となる．以下に HyperOmniVision の光学系と自由視点画像生成に必要なカメラパラメータのキャリブレーションについて説明する．

2.1 光学系

図2に HyperOmniVision 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ており，図2に示すように二つの焦点 O_M, O_C を持ち， O_C にカメラ部の光学中心が位置するようにミラーをカメラ部に取り付ける．参考文献12)より， Z 軸を鉛直軸とする3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 ，およびカメラ部の光学中心 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)．

$$\begin{aligned} \text{ミラー曲面} &: \frac{X^2 + Y^2}{a^2} - \frac{Z^2}{b^2} = -1 \\ \text{ミラー焦点 } O_M &: (0, 0, +c) \\ \text{カメラ部の光学中心 } O_C &: (0, 0, -c) \quad (c = \sqrt{a^2 + b^2}) \end{aligned} \quad (1)$$

O_C に光学中心を置いたカメラ部 (焦点距離 f) で全方位画像を撮影すると，双曲面の特性により，3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 (文献1) 参照)．

$$\begin{aligned} x &= \frac{Xf(b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}} \\ y &= \frac{Yf(b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}} \end{aligned} \quad (2)$$

2.2 キャリブレーション

環境中の3次元位置が既知の基準点 (以下，マーカと呼ぶ) を用いてカメラ部の画像中心 (C_x, C_y) と焦点距離 f ，世界座標系に対する位置 $T(t_x, t_y, t_z)$ ・姿勢 $R(r_x, r_y, r_z)$ の9パラメータを推定する．ここで推定した位置・姿勢情報は，後述の自由視点画像の生成結果に大きな影響を与え

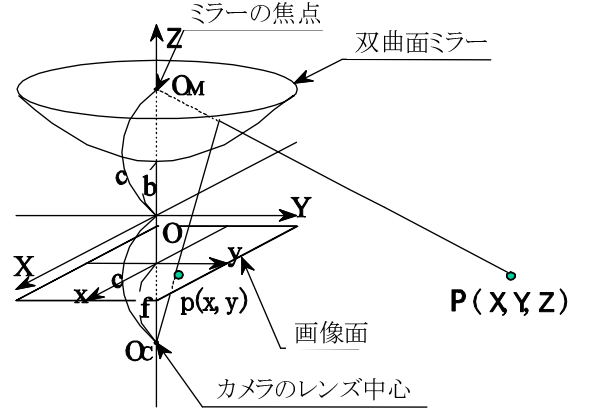


図2 HyperOmniVision の光学系
Camera model of HyperOmniVision.

るため，精度良く推定する必要がある．

本研究では，根岸ら¹³⁾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 ($C_x, C_y, f, t_x, t_y, t_z, r_x, r_y, r_z$) を推定する．その際，ミラーの内側の焦点 O_M はカメラ部の光軸上であり，カメラ部はミラーの外側の焦点 O_C に配置されていると仮定する．そして，ミラーの形状パラメータ a, b, c は既知とする．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，全方位カメラの焦点距離と世界座標系に対する位置・姿勢を推定する．

まず，画像中心 (C_x, C_y) 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画像中心を推定している．これに対し，本研究では HyperOmniVision 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しているため，画像中心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 $(x_i, y_i) (i = 1, \dots, N)$ ， r を円の半径とすると，これらの点と円方程式の誤差関数 X_c は以下の式で定義できる．

$$\begin{aligned} X_c &= \sum_{i=1}^N d_i^2 \\ d_i &= \sqrt{(x_i - C_x)^2 + (y_i - C_y)^2} - r \end{aligned} \quad (3)$$

式(3)の最小化によって (C_x, C_y) を推定する．

次に，上で推定した画像中心 (C_x, C_y) と暫定的に設定した7パラメータ ($f, t_x, t_y, t_z, r_x, r_y, r_z$) で3次元位置が既知のマーカが式(2)により投影される画像上の点と，実際に画像上で検出された点との2乗誤差を最小化することで，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

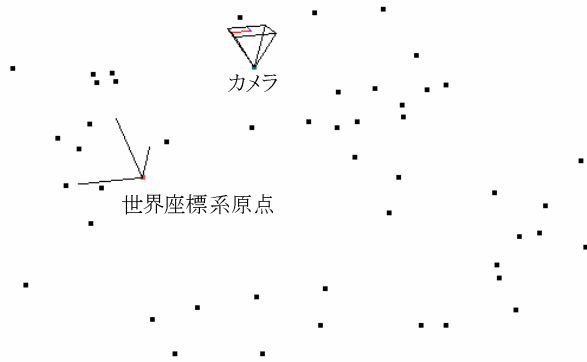


図 3 環境中のマーカと推定されたカメラの位置・姿勢
Markers and estimated camera position and posture.

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3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．

3. 多地点全方位画像からの自由視点画像生成

3.1 概要

環境中に配置された多地点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成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，平面的な視点移動を可能にすることや，生成した自由視点画像の品質を高くするためである．また，撮影環境の大部分は静止環境であり，動物体は全方位画像中の小領域であると仮定する．加えて，各全方位カメラは固定して設置されており，同期済みであるとする．その固定された各全方位カメラの位置・姿勢は2章で述べた方法により推定する．

図4に自由視点画像生成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¹⁴⁾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Morphing¹⁵⁾により自由視点画像を

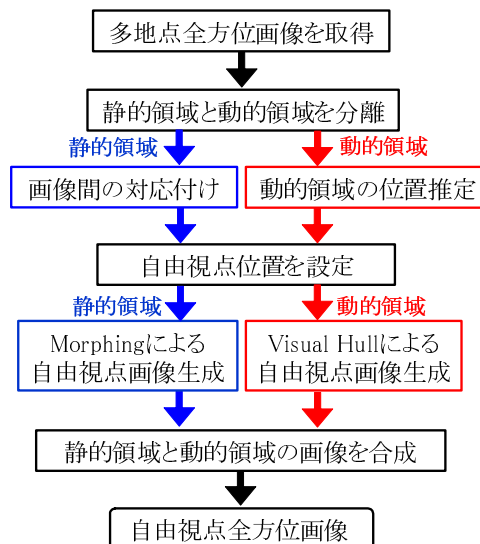


図 4 自由視点画像生成処理の流れ
Flow diagram of novel view generation.

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 Hull¹⁶⁾による Visual Hullの計算を行う．さらに，動物体の位置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生成の後，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，各処理について説明する．

3.2 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のは困難である．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¹⁴⁾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，それ以外の場合は動的領域と見做す．画像中の各背景画素の輝度 I は式(4)でモデル化される．

$$I = \bar{I} + \sigma \sin(2\pi\omega t) + k\xi \quad (4)$$

\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 $-1 \sim +1$ 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．

$$\bar{I} - \sigma - \xi \leq I \leq \bar{I} + \sigma + \xi \quad (5)$$

\bar{I} と σ 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， ξ 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 (5) 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．

3.3 静的領域に対する自由視点画像生成

前項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¹⁵⁾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ので，予め対応点を与えておくことができる．

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 5 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¹⁸⁾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．

実装上の工夫として，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．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，後述する合成処理のために PC のメインメモリ上に戻さず GPU のフレームバッファ上に保持しておく．

3.4 動的領域に対する自由視点画像生成

3.2 項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

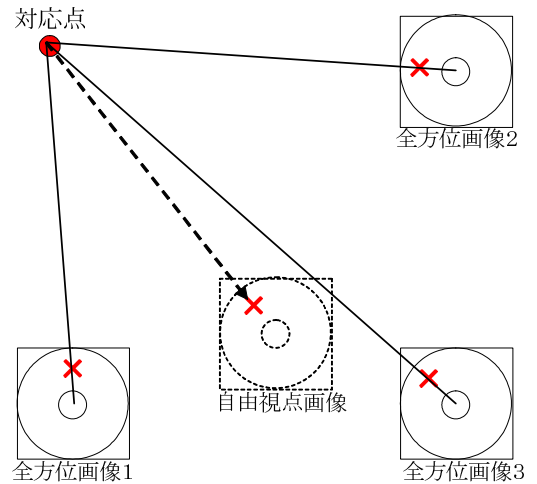


図 5 対応点の自由視点画像への投影
Projection of corresponding points to a novel view image.

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¹⁹⁾²⁰⁾．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 Hull¹⁶⁾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 Hull¹⁶⁾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，自由視点からの直線は Visual Hull を貫通する．交差線分がないならば Visual Hull を貫通しないと判断できる (図 6 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，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．
- 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 Hull¹⁶⁾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

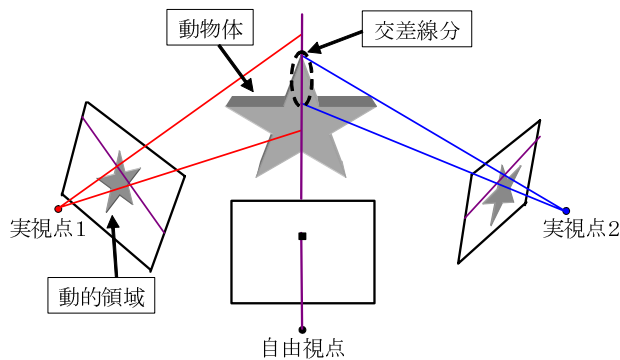


図 6 Visual Hull の貫通の判定
Estimation of penetration of visual hull.

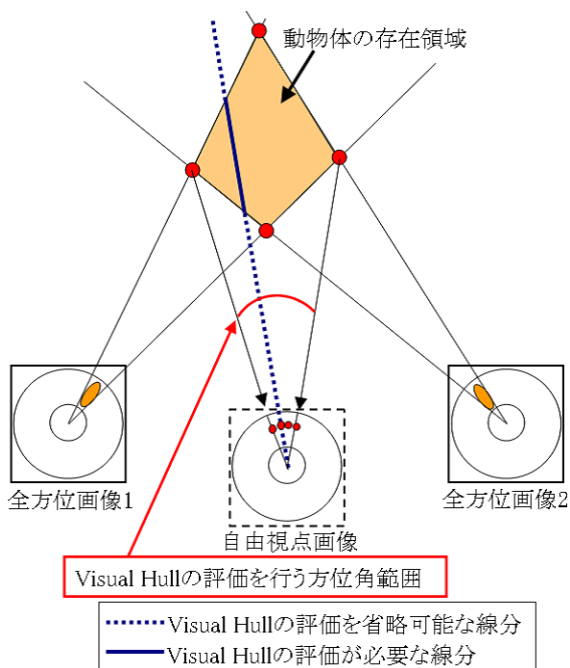


図 7 推定された動物体の存在領域
Estimated object existing region.

ために予め動物体の環境中での存在領域を推定し、その推定した領域を基に Visual Hull の計算が必要な自由視点画像上の画素を限定する。

本研究では、森田らの手法¹⁷⁾を用いて動物体が存在する領域を推定する。推定の過程で動物体が存在する領域を構成する直線の交点(図7参照)が求まる。この交点を式(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、自由視点からの直線が Visual Hull を貫通するか否かを評価する際、エピポラ線上の推定された動物体の存在領域以外では Visual Hull の評価を省略する(図7参照)。

実装上、動的領域の自由視点画像は PC のメインメモリー上に保持されている。後の合成処理を高速に行うために、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(非透過度)として 0 を設定する。

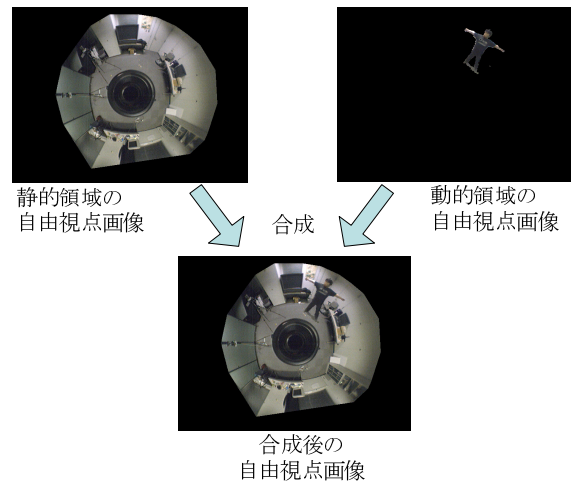


図 8 静的領域と動的領域の自由視点画像の合成
Synthesis of static and dynamic regions.

3.5 静的領域と動的領域の自由視点画像の合成

3.3, 3.4 項の処理により、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8参照)。

実装上の工夫として、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、フレームバッファに保持されている静的領域の自由視点画像の上に、アルファ値(非透過度)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。

4. 自由視点画像の定量的品質評価

本節では、シミュレーション画像を用いた自由視点画像の定量的品質評価を行う。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と、その環境中で人が自由視点位置から対話可能な距離に存在するような環境での使用を想定している。このような環境を 3DCG オブジェクトにより構築し、この環境中の 3 地点で描画した全方位画像を入力とし、ある自由視点における全方位画像を生成する。この自由視点全方位画像と実際にその視点位置で描画した全方位画像とで PSNR(Peak Signal-to-Noise Ratio) による定量的品質評価を行う。以下に、評価実験の条件と評価結果について述べる。

4.1 条件

評価実験で用いる全方位画像はシミュレーション環境中の 3 地点で取得する。各シミュレーション画像は、3DCG オブジェクトにより構成された環境を描画し生成されているので、全方位カメラの内部・外部パラメータは真値として用いることができる。全方位カメラの内部パラメータとして、焦点距離 f を 900.0 画素、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 とした。これらの値は、後述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。外部パラメータについては、環境中の

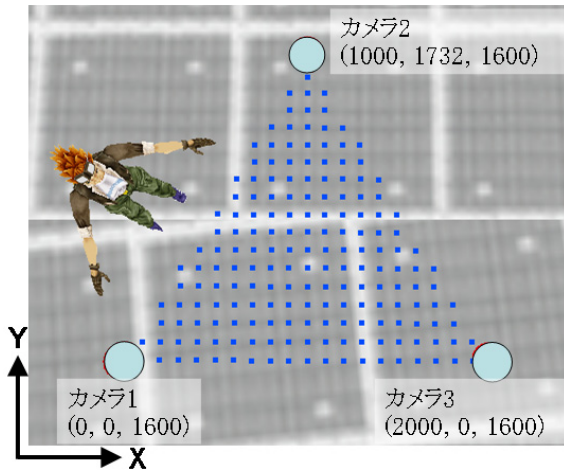


図 9 シミュレーション環境中のカメラ配置と自由視点位置
Camera positions and novel view-points in a simulated environment.

地面と平行な面を XY 座標，高さ方向を Z 座標とし，図 9 に示すような正三角形の頂点にカメラ（図中の大きな丸印）を配置した．図中の小さな点は自由視点位置を示し，動物体として手を広げた人物を配置した．本実験では，自由視点位置は入力画像が撮影された 3 地点からなる三角形内の 184 箇所とし，各視点位置で自由視点全方位画像とその真値となる全方位画像を生成した．その際，各入力全方位画像の解像度は 512×512 とし，出力全方位画像も同様の解像度とした．図 10 に，自由視点画像生成のために入力した全方位画像と対応点を示す．図 10 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 \times 印は対応点を示している．ここでは手動で約 200 点の対応点を与えた．図 10 下段は動的領域のみが含まれる全方位画像である．

本実験では，定量的品質評価の指標として先に述べたように PSNR を用いる．各自由視点位置で自由視点全方位画像とその真値となる全方位画像から PSNR を計算し，画像が真値からどの程度劣化しているかを評価する．

4.2 品質評価

図 11 に，各自由視点位置における PSNR の値を輝度マップとして示す．図中の濃度値はその位置で生成された自由視点画像の品質を表し，明るいほど生成された自由視点画像の品質が高く，暗いほど自由視点画像の品質が低いことを表す．PSNR の最大値は $46.33[\text{dB}]$ ，最小値は $19.26[\text{dB}]$ ，平均値は $21.63[\text{dB}]$ であった．図 11 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，自由視点位置が入力画像が撮影された位置と同じ場合には，それぞれの画像の差はほとんど感じることはできなかった．生成画像の品質が低い位置は，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

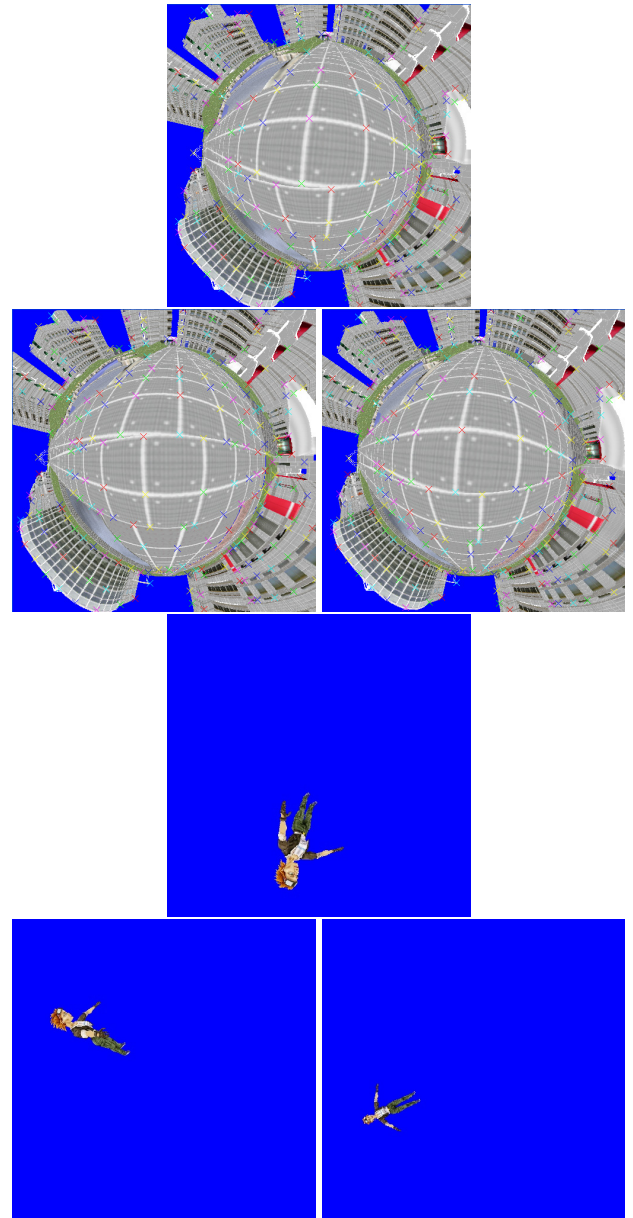


図 10 評価実験に用いた静的領域の画像とその対応点（上段）および動的領域の全方位画像（下段）

Input data for evaluation: (top) images including static region and corresponding points and (bottom) images including a dynamic region.

が低くなっている．図 12 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．

本研究で想定するアプリケーションでは，自由な視点位置で全方位を観賞可能であることを要求しており，本評価結果からその要求を満たす画像が生成できていると言える．ただし，環境中の動物体については，その動物体が複雑な形状であり Visual Hull で表現が困難な場合には提示画像の品質は低くなる．

5. 自由視点テレプレゼンスシステム

3 節で述べた自由視点画像生成手法により，利用者に自由視点画像を実時間提示可能なテレプレゼンスシステムのプロトタイプを構築した．本システムは，遠隔地で撮影さ

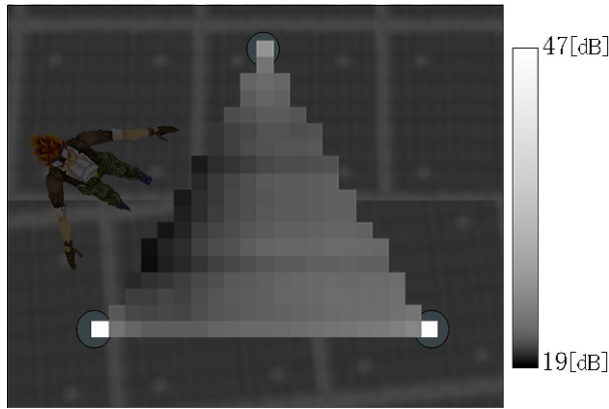


図 11 PSNR の輝度マップ表示
Luminance map of PSNR evaluation.

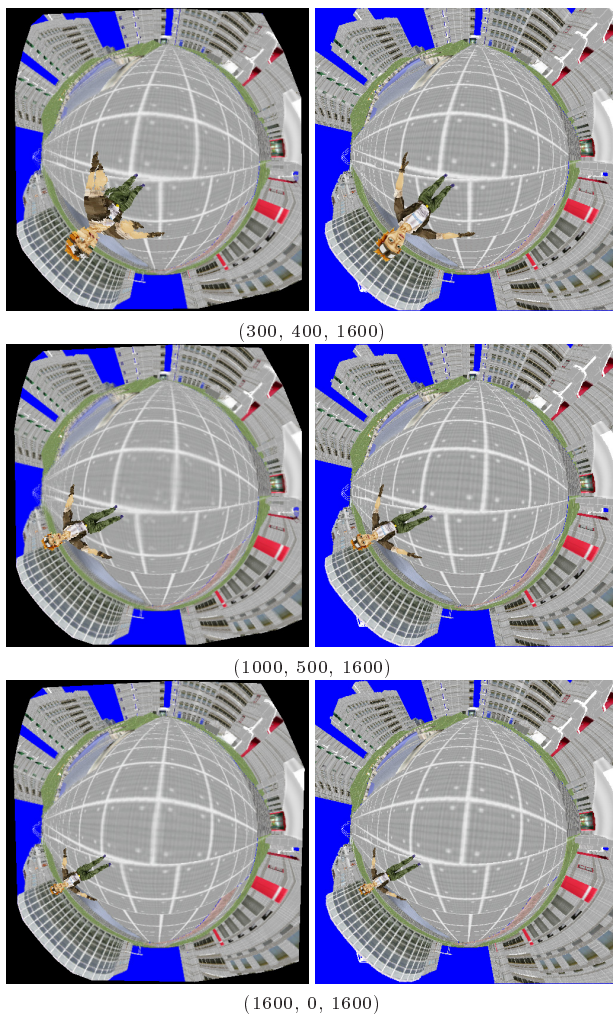


図 12 生成した自由視点全方位画像 (左) と真値の全方位画像 (右)
Examples of novel view images: (left) generated novel view images and (right) grand truth images.

れた複数の全方位ライブ映像をネットワークを介して伝送し、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臨場感高く観賞可能である。以下に、システムの構成と処理の流れおよび実験結果について詳述する。

5.1 システム構成

図 13 にプロトタイプシステムの構成を示す。本システム

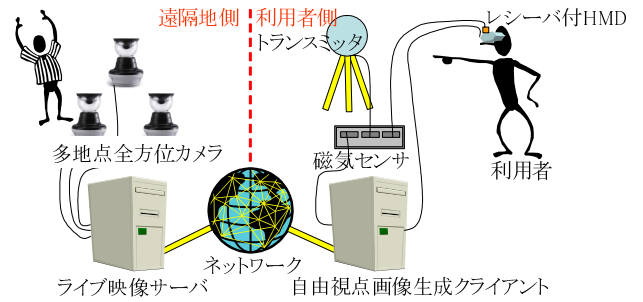


図 13 プロトタイプシステムの構成
Configuration of prototype system.

表 1 プロトタイプシステムの機器構成
Components of prototype system.

全方位カメラ	解像度：640x480， 最大フレームレート：15fps， 画角：水平 360 度，垂直 62 度
ライブ映像サーバ	CPU:AMD Athlon64 X2 4400+， メモリー：2GB
自由視点画像生成クライアント	CPU：Intel Pentium4 3.2GHz， メモリー：2GB， GPU：nVidia GeForce6600GT

は、遠隔地に配置した全方位カメラの画像を送信する PC (ライブ映像サーバ) と、利用者側で映像を受信し提示画像を生成する 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システム構成を詳述する。

[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送信する。

[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、三つの全方位画像とそれらに対応する三つのマスク画像 (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磁気センサ (Polhemus 社製 3SPACE FASTRAK) 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、磁気トランスミッタにはロングレンジャー (図 14 左) を使用する。システムの利用者は、図 14 右に示す磁気レシーバ付き 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、提示画像を観賞する。

表 1 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。

5.2 処理の流れ

本システムでは、ライブ映像サーバと自由視点画像生成



図 14 磁気センサ(左)と磁気レシーバ付き HMD(右)
Magnetic sensor (left) and receiver attached HMD (right).

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 15 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 15 に従って説明する。

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]

- ・ 環境中に複数配置された全方位カメラのキャリブレーションを 2.2 項で述べた方法により行う。
- ・ 全方位画像中の画像の 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。
- ・ 3.2 項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]

- ・ 3.3 項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 項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(光線ベクトル)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。

オンライン処理では、時々刻々と取得する全方位動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]

- ・ 多地点全方位画像を取得し、背景差分処理を行い、その結果を自由視点画像生成クライアントに送信する。ここでライブ映像サーバが送信するのは、取得した全方位画像そのものと背景差分処理の結果得られた動的領

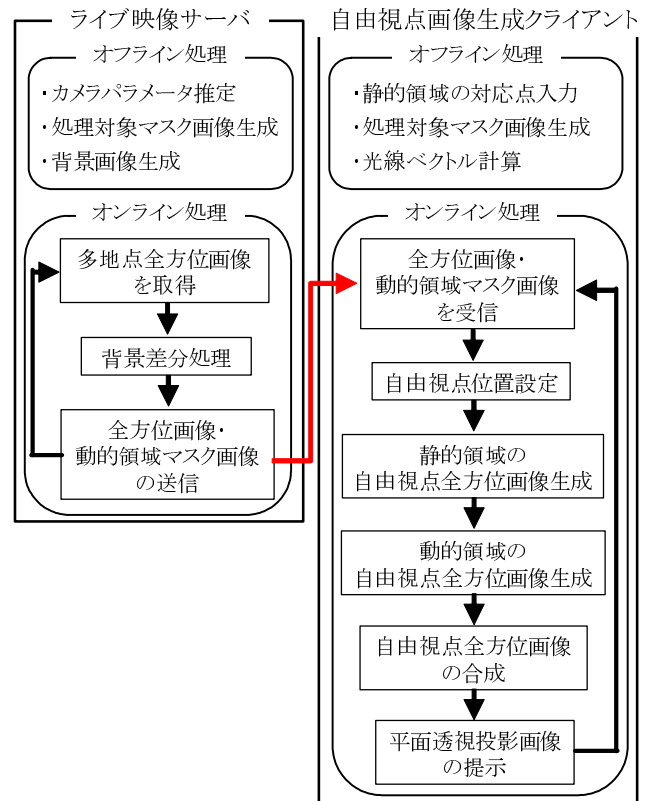


図 15 プロトタイプシステムの処理の流れ
Process flow of prototype system.

域のマスク画像(2 値画像)である。これにより、静的領域と動的領域が分離されたカラー画像を送るよりも伝送データ量の削減が可能である。

[自由視点画像生成クライアント]

- ・ 全方位画像と動的領域のマスク画像を受信し、磁気センサから利用者の視線方向・視点位置を取得した後、3 節で述べた自由視点画像生成処理を行う。この際、3.4 項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。
- ・ 自由視点全方位画像を生成後、平面透視投影変換を行う。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¹⁾を行う。
- ・ 生成された平面透視投影画像を利用者が装着した HMD に提示する。

以上の処理を繰り返し行い、1 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。

5.3 実験結果

図 16 上に示すように、室内環境の 3 地点に全方位カメラを配置して、テレプレゼンス実験を行った。オフライン処理で推定されたカメラパラメータを基に、各カメラの位置関係を図 16 下に示す。また、各カメラのパラメータの詳細を表 2 に示す。本実験では、全方位画像中の静的領域の



図 16 カメラ配置と推定されたカメラ位置
Camera arrangement and estimated positions.

表 2 実験に用いたカメラパラメータ
Camera parameters.

ミラーの形状 (各カメラ共通)	$a = 26.26[\text{mm}]$ $b = 35.49[\text{mm}]$
カメラ 1	位置 $(-937, -819, 1542)$ [mm] 姿勢 $(-0.258, 0.200, -0.586)$ [度] 焦点距離 886[pixel]
カメラ 2	位置 $(-1880, -2054, 1472)$ [mm] 姿勢 $(-2.208, -0.810, -7.669)$ [度] 焦点距離 878[pixel]
カメラ 3	位置 $(93, -1981, 1553)$ [mm] 姿勢 $(-0.050, -3.299, -7.932)$ [度] 焦点距離 895[pixel]

対応点を 78 点手動で与えた。

図 17 に、実験環境の様子と自由視点における生成画像を示す。利用者の視点位置・視線方向に応じた画像が生成できていることが確認できる。また、動的領域に対しては利用者が見ている範囲のみを計算対象とするため、自由視点全方位画像中には視線方向にいる人物以外は映っていないことが確認できる。このとき、利用者に提示される映像の各フレームは約 200ms (約 5fps) で生成された。内訳は、全方位画像とマスク画像の受信に約 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 となっていた。それに伴い、利用者の視線方向・視点位置の取得からその視線・視点に対応する画像が提示されるまでの時間遅延は約 120ms となる。また、ライブ映像サーバでの全方位画像取得から利用者への画像提示

までの時間遅延は約 320ms であった。その内訳は、ライブ映像サーバでの撮影からその画像の伝送開始までの待機時間 (自由視点画像生成クライアントでの画像受信以降の処理時間) が約 120ms、自由視点画像生成クライアントが画像を受信開始してから画像を提示するまでの処理時間が約 200ms となっていた。

利用者に提示された画像は、評価実験と同様に、利用者の自由視点位置がカメラに近いほど、良好な結果となった。しかし、本実験では動物体がほぼ直立状態であったため、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違和感を感じることはなかった。評価実験の静的環境は、撮影対象とカメラとの距離が長いい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 (図 18 左参照)。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 (図 18 右参照)。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。また、静的領域と動的領域を別々に生成し合成された自由視点画像であるが、静的領域と動的領域間で位置のずれはほとんど感じられなかった。これは、本手法で用いている自由視点画像生成手法が IBR に基づいていながら、幾何学的な情報も持ち合わせているからであると考えられる。

6. む す び

本論文では、利用者が遠隔地の動的環境中をライブかつ自由な視線方向・視点位置で観賞可能なテレプレゼンスを提案した。そのために、以下の要求機能 (A) ~ (D) を実現した、

- (A) 動的な撮影環境全体に対応：配置されたカメラ群の内側・外側の動的な環境全体を多地点の全方位カメラによって撮影することにより実現。
- (B) 遠隔地映像の伝送：多地点全方位画像をネットワークを介して伝送することで実現。
- (C) ライブ映像からの自由視点画像生成：全方位画像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 IBR に基づく自由視点画像生成手法を用い、さらに計算量削減手法を併用することで高速に自由視点画像を生成。
- (D) 利用者の動作に応じた画像提示：位置・姿勢センサと HMD を用いて実現。



(a) 遠隔地の様子

(b) 利用者の様子

(c) 自由視点全方位画像

(d) 平面透視投影画像

図 17 テレプレゼンスの様子と生成された自由視点画像
Result of novel view telepresence and experimental environment.



図 18 対応点が与えられていない部分の歪み (左) と半透明に映りこむカメラ (右)

Distortion caused by no corresponding points (left) and semitranslucent cameras (right).

生成画像の定量的な品質評価として、シミュレーション画像を用いて自由視点画像生成を行い、その画像の PSNR を算出した。実際にプロトタイプシステムを構築し、遠隔地の動的環境を撮影した全方位ライブ映像をネットワークを介して伝送し、その映像から利用者に提示する自由視点画像を生成した。利用者は、遠隔地に居るかのように環境中を実際に動き、その動作に応じた画像を観賞することで高い没入感覚を与えられた。このことから、提案したテレプレゼンスは、観光地の多地点に全方位カメラを配置し、その環境を遠隔地で体験するような仮想旅行のようなアプリケーションに応用できると考える。ただし、位置・姿勢センサ付き HMD を装着した状態で広範囲を歩き回るのは現実的ではないので、ジョイスティック等のインタフェースの利用も検討すべきである。また、遠隔監視やビデオ会議システム等に応用することで、遠隔地の状況を空間的に把握することも可能となる。

今後の課題としては、自由視点画像生成の品質向上やより広範囲をウォークスルー可能なシステムへの拡張などが考えられる。自由視点画像の品質改善のためには、静的領域の対応点を密に与える方法の検討の他、前節で述べた相互のカメラの映りこみを軽減する方法の適用を検討する。また、高解像度の全方位カメラを用いることで、品質の向上が可能である。プロトタイプシステムをより広範囲をウォークスルー可能なシステムへ拡張するためには、自由視点位置に応じて使用するカメラを動的に変化させ、使用するカメラが切り替わる際の違和感を低減する必要があると考えられる。

〔文 献〕

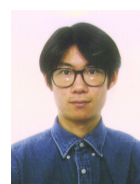
- 1) Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya: "Telepresence by real-time view-dependent image generation from omnidirectional video streams", *Computer Vision and Image Understanding*, **71**, 2, pp.154-165 (1998)
- 2) 池田聖, 佐藤智和, 横矢直和: "全方位型マルチカメラシステムを用いた高解像度な全天球パノラマ動画像の生成とテレプレゼンスへの応用", *日本バーチャリアリティ学論*, **8**, 4, pp.443-450 (2003)
- 3) 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", *信学論*, **J88-D-II**, 8, pp.1750-1753 (2005)
- 4) S. E. Chen: "Quicktime VR: An image-based approach to virtual environment navigation", *Proc. of SIGGRAPH '95*, pp.29-38

(1995)

- 5) T. Kanade, P. Rander, and P. J. Narayanan: "Virtualized reality: Constructing virtual worlds from real scenes", *IEEE Multimedia*, **4**, 1, pp.34-47 (1997)
- 6) 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*, **5**, 3, pp.303-316 (2003)
- 7) M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Szeliski and R. Hartley: "Image-based interactive exploration of real-world environments", *IEEE Computer Graphics and Applications*, **24**, 3, pp.52-63 (2004)
- 8) T. Koyama, I. Kitahara, and Y. Ohta: "Live mixed-reality 3D video in soccer stadium", *Proc. of IEEE / ACM Int. Symp. on Mixed and Augmented Reality (ISMAR03)*, pp.178-187 (2003)
- 9) 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観賞システム", *映情学誌*, **58**, 4, pp.529-539 (2004)
- 10) D. G. Aliaga and I. Carlbom: "Plenoptic stitching: A scalable method for reconstructing 3D interactive walkthroughs", *Proc. of SIGGRAPH2001*, pp.443-451 (2001)
- 11) 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*, **23**, 3, pp.600-608 (2004)
- 12) 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系 HyperOmni Vision の提案", *信学論*, **J79-D-II**, 5, pp.698-707 (1996)
- 13) 根岸善朗, 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, 2P2-L1-49, (2004)
- 14) K. Yamazawa and N. Yokoya: "Detecting moving objects from omni-directional dynamic images based on adaptive background", *Proc. of 10th IEEE Int. Conf. on Image Processing (ICIP2003)*, **III**, pp.953-956 (2003)
- 15) 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*, **II**, pp.353-358 (2002)
- 16) W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan: "Image-based visual hulls", *Proc. of SIGGRAPH2000*, pp.369-374 (2000)
- 17) 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", *信学論*, **J88-D-II**, 5, pp.864-875 (2005)
- 18) P. Heckbert Ed.: "Graphics Gems IV", Academic Press, pp.47-59 (1994)
- 19) 小川軍, 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", *情処学論: コンピュータビジョンとイメージメディア*, **42**, SIG6(CVIM2), pp.33-43 (2001)
- 20) 上田恵, 有田大作, 谷口倫一郎: "多視点動画像処理による 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-", *情処学論*, **46**, 11, pp.2768-2778 (2005)



いしかわ ともや
石川 智也 2003 年, 岡山理科大学・工学部・情報工学科卒業。2005 年, 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在, 同大博士後期課程に在学中。複合現実感の研究に従事。学生会員。



やまざわ かずまさ
山澤 一誠 1992 年, 大阪大学・基礎工学部・情報工学科卒業。1996 年, 同大大学院博士後期課程中退。同年, 奈良先端科学技術大学院大学情報科学研究科助手。2002 年, 同大助教授, 現在に至る。コンピュータビジョン, 複合現実感の研究に従事。博士 (工学)。



よこや なおかず
横矢 直和 1974 年, 大阪大学・基礎工学部・情報工学科卒業。1979 年, 同大大学院博士後期課程修了。同年, 電子技術総合研究所入所。以来, 画像処理ソフトウェア, 画像データベース, コンピュータビジョンの研究に従事。1986~87 年, マッギル大学・知能機械研究センター客員教授。1992 年, 奈良先端科学技術大学院大学教授。工学博士。正会員。