

空撮全方位動画像を用いた蓄積再生型拡張テレプレゼンス における幾何学的・光学的整合性の実現

大倉 史生[†] 神原 誠之[†] 横矢 直和[†]

[†] 奈良先端科学技術大学院大学 〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{fumio-o,kanbara,yokoya}@is.naist.jp

あらまし 本論文では、実環境の情景に仮想物体を重畳する拡張現実感 (Augmented Reality: AR) とユーザに没入感の高い遠隔地の情景を提示するテレプレゼンスを組み合わせた拡張テレプレゼンスにおいて、特にユーザがオフラインで情景を体験する用途に応用可能な、あらかじめ撮影・蓄積された動画像を用いた蓄積再生型拡張テレプレゼンスシステムの技術課題である幾何学的・光学的整合性の実現手法について述べる。本研究では無人飛行船から撮影された空撮動画像を用い、幾何学的整合性を実現するためのカメラ位置・姿勢推定、および光学的整合性を実現するための欠損のない環境マップ推定とグローバルイルミネーションレンダリングを、それぞれオフラインで高品質に実現する。最後に、実際に平城宮跡上空から撮影された動画像を用いた AR シーンの生成実験および平城宮跡で行われた一般公開実験について述べ、さらに定量評価を通じて提案システムの有効性を検証する。

キーワード 拡張テレプレゼンス, 空撮, 無人飛行船, 幾何学的整合性, 光学的整合性

1. はじめに

近年、ユーザに没入感の高い遠隔地の情景を提示するテレプレゼンス [1] に関する研究が盛んに行われている。遠隔地の情景を実時間でユーザに提示する実時間テレプレゼンスシステムは遠隔コミュニケーションや遠隔手術、災害地等の監視等に用いられている。一方で、あらかじめ蓄積された遠隔地の情景を用いる蓄積再生型テレプレゼンスシステムとして、地図やウォークスルー、エンタテインメント等を応用とする様々な研究が行われている。特に全方位カメラを用いた見回し可能な蓄積再生型テレプレゼンスは、近年では Google 社の Google Street View などとして広く普及している。

さらに、遠隔地の情景をそのまま提示するだけでなく、その場所に関連する情報の提示を行うために、カメラ等で取得された実環境の情景に仮想物体を重畳する拡張現実感 (Augmented Reality: AR) [2] とテレプレゼンスを組み合わせることで、ユーザへの情報提示を効果的に行う研究が行われている [3-5]。本研究ではこれを拡張テレプレゼンス (Augmented Telepresence) と呼ぶ。特に、空撮動画像を用いた拡張テレプレゼンスシステム [4,5] は地上での撮影と比較して広範囲の情景を取得できるという特徴を持ち、ヘリコプタ等の移動体の遠隔操縦支援や景観シミュレーション等を目的とした実時間システムについて研究が行われており、さらなる応用が期待できる。

本研究では、空撮動画像を用いた拡張テレプレゼンスによる「時空を超えた」バーチャル歴史観光システムを構築する。2010 年は奈良にかつて存在した平城京が設立されて 1300 周年であり、平城遷都 1300 年祭が開催された。提案システムは、奈良時代の平城宮に存在した建造

物を復元した「時間を超える」CG を仮想物体とし、無人飛行船から撮影・蓄積された「空間を超える」動画像に重畳する。本研究では蓄積された動画像を用いた拡張テレプレゼンスを蓄積再生型拡張テレプレゼンスと呼ぶ。従来、蓄積再生型拡張テレプレゼンスについては地上の 1 地点から撮影された全方位画像を対象として GIS 情報の可視化等を目的とした研究 [6,7] が行われているが、広域屋外環境における移動撮影を意図しておらず、幾何学的整合性または光学的整合性を考慮していなかった。本研究で実現するバーチャル歴史観光システムは、従来のシステムと比較しリアリティおよび没入感の高い体験をユーザに提供するために幾何学的・光学的整合性を実現する。これらのアプローチは、本研究のようなバーチャル観光旅行等のエンタテインメント用途のみならず、同様に高品質な仮想物体を重畳し、高い没入感を提供することが望まれる景観シミュレーションや都市計画等にも適用が可能であり、近年広く普及している全方位画像を用いた蓄積再生型テレプレゼンスによるアプリケーションにも容易に応用できる。

全方位動画像を用いた蓄積再生型拡張テレプレゼンスを実現するため技術課題および解決のためのアプローチを、幾何学的・光学的整合性の観点で以下に概説する。

(1) 幾何学的整合性の実現

幾何学的整合性における最も重要な課題である仮想物体の位置合わせのために動画像と GPS を用いたカメラ位置・姿勢推定手法 [8] を適用する。広域屋外環境に適用可能な動画像からのカメラ位置・姿勢推定手法として Structure-from-Motion (SfM) が挙げられるが、一般的に SfM では絶対的なスケールが得られないため、GPS と併用することで既知の三次元点座標を測量する等の事

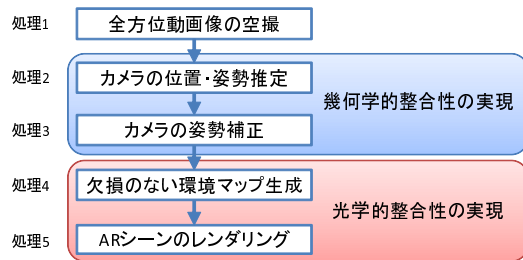


図1 提案システムの処理の流れ

前準備を行わずにスケール情報付きのカメラ位置・姿勢を得る。また、SfMを用いたカメラ位置姿勢推定において、センサの誤差やトラッキング誤差に起因する推定誤差が発生する。本研究ではシーンに遠方の物体が多いため、特に姿勢方向の誤差が問題となる。そのため、まず推定されたカメラの位置・姿勢情報を用いて全方位動画の姿勢を統一し、姿勢統一された動画を用いて姿勢成分の誤差の補正を行う。

(2) 光学的整合性の実現

実環境の全方位画像を光源環境マップとして用い、仮想物体に対してImage Based Lighting (IBL)を用いたグローバルイルミネーション (GI) レンダリングを行うことで実物体と仮想物体間の光源環境を統一する。このとき、実環境を撮影した全方位画像には、飛行船の映り込みや全方位カメラで撮影できない部分による死角領域が存在する。そのため、死角領域を補完することで情景の欠損がない全方位動画画像を生成する。動画画像の全フレームから情景を推定し、動画画像全フレームを通じて死角となる部分では天空光モデルを用いて天空の輝度分布を推定することで、死角領域の補完を行う。

本論文では、2章で飛行船と全方位カメラを用いた空撮システムおよび幾何学的・光学的整合性の実現手法について述べる。次に3章では平城宮跡上空から撮影された動画画像を用いた全方位ARシーンの生成実験とその結果、および平城宮跡で行われた提案システムの一般公開実験のための拡張テレプレゼンスシステムの構成と公開実験の様子について述べ、さらに定量評価による提案システムの性能の検証を行う。

2. 飛行船からの空撮全方位動画画像を用いた拡張テレプレゼンス

2.1 システムの処理概要

提案システムの処理の流れを図1に示す。以下に、各処理を順に概説する。

(処理1) 空撮動画を無人飛行船に搭載された全方位カメラにより撮影し、見回し可能なテレプレゼンスシステムを実現する。

(処理2) GPS、姿勢センサ、および空撮動画画像を用いてカメラの位置・姿勢を推定する。

(処理3) カメラの位置・姿勢推定処理において発生するセンサの誤差やトラッキング誤差に起因する推定誤差を補正するため、姿勢統一された全方位動画画像を用いて姿

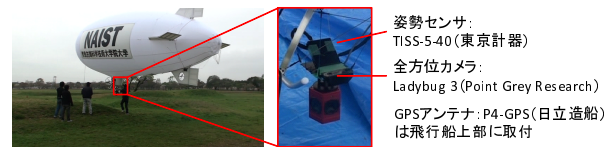


図2 無人飛行船の外観と搭載機器

勢成分の誤差の補正を行う。

(処理4) 本研究では飛行船の全方位動画画像への映り込みや全方位カメラで撮影できない部分を死角領域と呼び、これを消去・補完することで欠損のない環境マップを生成する。

(処理5) 最後に、欠損のない環境マップを光源として市販のレンダリングソフトウェアにより仮想物体にグローバルイルミネーション (GI) を適用しレンダリングを行う。

以下、全方位動画画像の空撮、幾何学的・光学的整合性の実現についての詳細を述べる。

2.2 全方位動画画像の空撮

提案システムでは、図2に示す全長12mのラジコン無人飛行船を用いて空撮を行う。飛行船には空撮動画画像を取得するための全方位マルチカメラシステム Ladybug3 (Point Grey Research社製) を下向きに取り付け、離散的な位置情報取得のためのGPSとして電子基準点網によるネットワーク補正を利用するDifferential GPSであるP4-GPS (日立造船株式会社製)、姿勢センサとして光ファイバージャイロ TISS-5-40 (東京計器株式会社製) を搭載する。全方位カメラによって取得された画像は、全方位カメラにおける各カメラの内部パラメータおよびカメラ間の関係を用いて図3に示すようなパノラマ画像に変換される [9]。各センサの詳細を表1に示す。飛行船に搭載された全方位カメラおよびセンサは、データ蓄積のために1台のノートPCに接続される。

2.3 カメラの位置・姿勢推定

提案システムでは、動画画像からカメラ位置・姿勢を推定するStructure-from-Motion (SfM) アルゴリズムとGPSの位置情報を併用するハイブリッド手法 [8] を用いてカメラ位置・姿勢を推定する。一般に、動画画像を用いた逐次的なトラッキングには蓄積誤差が発生する。SfMアルゴリズムは全体最適化を行い誤差を動画画像全体に分

表1 搭載センサの仕様

センサ	モデル	詳細
全方位マルチカメラシステム	Point Grey Research Ladybug3	5400 × 2700 画素 15 fps 搭載カメラ: 6 個
GPS	日立造船 P4-GPS	Differential GPS 4Hz 水平 RMS 40.4cm 垂直 RMS 53.5cm
姿勢センサ	東京計器 TISS-5-40	精度: 3°/h 500Hz



図3 全方位カメラで得られる空撮全方位画像

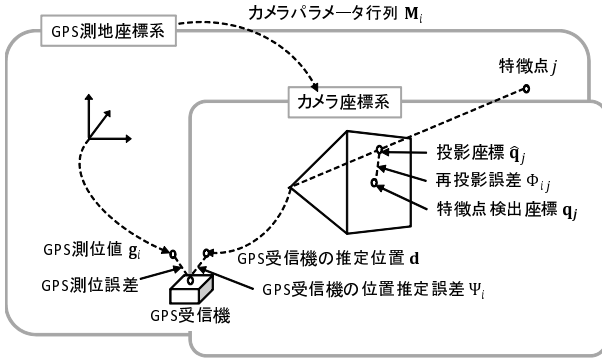


図4 カメラ位置・姿勢推定 [8] における座標系，記号および誤差の定義

散させることによって蓄積誤差を抑制しているが，原理上，蓄積誤差に起因する動画像全体の誤差を完全に解消することは難しい．また，SfM アルゴリズムでは相対的なカメラ位置・姿勢しか得られずスケールが未知である．これに対して GPS の位置情報には蓄積誤差が発生せず，絶対座標系である GPS 測地座標系で測位が行われスケールが既知である．文献 [8] では，それらを併用することで，絶対座標系上でのカメラ位置・姿勢推定を行う．カメラ位置・姿勢推定手法 [8] の概要を以下に示す．

カメラ位置・姿勢推定に用いる座標系，記号および誤差の定義を図 4 に示す．一般に SfM アルゴリズムでは第 i フレームにおける画像上で検出された特徴点 j の座標 \mathbf{q}_{ij} ，および推定された特徴点 j の三次元位置を再投影することで計算される画像上への投影座標 $\hat{\mathbf{q}}_{ij}$ を用いて，第 i フレームにおける特徴点 j の再投影誤差関数 Φ_{ij} を以下のように定義する．

$$\Phi_{ij} = |\mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij}|^2. \quad (1)$$

本研究の枠組みにおいては，GPS 受信機を用いるために第 i フレームにおける GPS 測地座標系上での GPS 受信機の位置を \mathbf{g}_i ，カメラパラメータ行列を \mathbf{M}_i とするとき，カメラに対して GPS 受信機は相対的に固定されているため，カメラ座標系上での GPS 受信機の位置 \mathbf{d} は一定となり，以下の関係が成立する．

$$\mathbf{M}_i \mathbf{g}_i = \mathbf{d}. \quad (2)$$

しかし，一般に \mathbf{M}_i には推定誤差が含まれるため，GPS 受信機の位置推定誤差関数 Ψ_i を以下のように定義する．

$$\Psi_i = |\mathbf{d} - \mathbf{M}_i \mathbf{g}_i|^2. \quad (3)$$

誤差関数 Φ_{ij} および Ψ_i を用い，評価関数 E は以下のように定義される．

$$E = \frac{\omega}{|\mathbf{F}_g|} \sum_{i \in \mathbf{F}_g} \Psi_i + \frac{(1-\omega)}{\sum_i |\mathbf{S}_{p_i}|} \sum_i \sum_{j \in \mathbf{S}_{p_i}} \Phi_{ij}. \quad (4)$$

ここで， ω は重み係数， \mathbf{S}_{p_i} は第 i フレームにおいて観測された特徴点の集合， \mathbf{F}_g は GPS 測位データが観測されたフレーム集合を表す．カメラ位置・姿勢 \mathbf{M}_i は，適切な ω を用いて評価関数 E を勾配法などにより最小化することで求められる．

本研究では，全方位カメラおよび GPS 受信機は飛行船船体に固定されているものとし，カメラ座標系上での GPS 受信機の位置 \mathbf{d} は飛行船の直径およびゴンドラの高さから計算する．また，カメラで撮影された動画像の各フレームおよび GPS 受信機の測位についてのタイムスタンプを保持しておき，GPS 測位のタイムスタンプと最も近いタイムスタンプを持つフレームに GPS 測位データを付与する．

文献 [8] においては位置推定に用いる GPS として RTK-GPS を用いているが，本研究では，飛行船のペイロードの制限から軽量の Differential GPS を用いる．RTK-GPS の計測精度はおおよそ数 cm から数十 cm 程度であり，Differential GPS と比較して精度が高い．本研究では，文献 [8] で用いられる $\omega = 10^{-9}$ と比較し小さな重み $\omega = 10^{-13}$ を用いることで，GPS の計測誤差による影響を低減する．また，本研究において大量のフレームを一度に扱う場合，文献 [8] の手法を用いると，GPS の測位誤差に起因して E が増大した結果，正しく推定が行われない場合があり，一度に全フレームの推定を行うのは困難である．そこで，本研究では全フレームに対し一度にカメラ位置・姿勢の推定を行わず，数百フレームの短い動画像に分割して推定を行う．また，各動画像間におけるカメラ位置・姿勢推定値の連続性を実現するため，各動画像を重複させる．重複したフレームにおいては，位置は 2 つの推定値を線形補間し，姿勢は 2 つの推定値の四元数を球面線形補間することによって推定値を一意に求める．

評価関数の最小化結果は，局所解の存在が問題となり初期値に大きく依存する．一般に，前フレームのカメラ位置・姿勢が正しく推定された場合，連続するフレーム間でのカメラの移動量が微小であるとして，前フレームで推定されたカメラ位置・姿勢を次フレームのカメラ位置・姿勢の初期値とすることができる．しかし，動画像の 1 フレーム目については初期値を決めることが難しい．文献 [8] では，測量された実環境の 3 次元座標と初期フレームの画像座標の対応を与えることにより，1 フレーム目のカメラ位置・姿勢の初期値を計算している．しかし，本研究では広域にわたる測量が難しく，人的コスト

が大きくなることから、飛行船に搭載した姿勢センサによって得られた姿勢情報を1フレーム目のカメラ姿勢の初期値としている。

2.4 カメラの姿勢補正

本研究では2.3において全方位カメラの姿勢が得られているため、これを用いて全方位動画像の姿勢統一を行う。第*i*フレームにおけるカメラパラメータ行列 \mathbf{M}_i の回転成分を \mathbf{R}_i としたとき、全方位動画像を球面に投影し $\mathbf{R}_i^{-1} = \mathbf{R}_i^T$ によって変換を行うことでGPS測地座標系における球の姿勢を統一する。再びパノラマ画像に展開することで、姿勢が統一された全方位動画像を得る。カメラの姿勢補正処理は、姿勢統一後の動画像に対して行われる。

本研究では遠方のシーンを扱うために2.3で推定するカメラの位置・姿勢のうち姿勢誤差が顕著に現れる。姿勢誤差の原因は姿勢の初期値として与えられた姿勢センサの誤差や、GPSの測位誤差に起因して生じる誤差が挙げられる。さらに、仮想物体を重畳した際にはドリフト誤差と比較しジッタ誤差が大きな違和感の原因となる。カメラの姿勢のジッタ誤差は、姿勢を統一した動画像において情景の細かな揺れとなるため、姿勢が統一された動画像上の揺れを抑制することにより、推定されたカメラ位置・姿勢のうち姿勢成分のジッタ誤差を補正する。以下に、誤差補正処理を詳述する。

- (1) 姿勢変化に依存して画像上での位置が変化する遠方の特徴点を用いて、連続する2フレーム間のオプティカルフローを算出する。本研究では、実空間は平面に近いと仮定して、仰角10度未満かつ俯角10度未満程度の水平線付近に存在する特徴点を用いる。
- (2) 2フレーム間の対応点を球面上にマッピングし、2フレーム間の回転パラメータを推定する。回転パラメータはRoll-Pitch-Yawの3パラメータで表され、対応点間の距離の2乗和で定義されるエネルギー関数を非線形最小化問題を解き最小化することにより推定される。
- (3) 推定された回転の逆変換を行うことにより、2フレーム間の姿勢誤差によるジッタを補正する。
- (4) 上記の処理を動画像の初期フレームから繰り返すことで、動画像全体のジッタを補正する。

2.5 欠損のない環境マップの生成

提案システムで撮影される全方位画像には、死角領域、すなわち飛行船の映り込みや全方位カメラで撮影できない領域が含まれる。本研究ではその領域の消去・補完を行い、死角のない全方位動画像を生成することにより実環境の光源環境推定および死角領域の実環境の情景への遮蔽による没入感低下への対策を行う。このとき、カメラと飛行船は相対的に固定されているため、死角領域は動画像全体で一意に指定できる。そこで、姿勢統一前の全方位画像を用いて死角領域を示すマスクを作成し、各



(a) 飛行船の映り込みによる地上 (b) 地上領域に存在する死角領域への遮蔽。塗りつぶされた領域はの補完結果。死角領域を示す。

図6 地上領域に存在する死角領域の補完

フレームの姿勢を用いて2.4と同様の処理を行うことによって各フレームのマスクを生成しておく。

本研究において死角領域の大部分は空に存在するが、飛行船の姿勢変化により図6(a)のように死角領域が地上を遮蔽する場合がある。本研究では、全方位動画像における仰角5度以上に地上の構造物が映りこんでいないと仮定して空領域と定義し、それ以外を地上領域として推定を行う。以下、死角のない全方位動画像の生成処理の流れを示し、各処理の詳細を述べる。

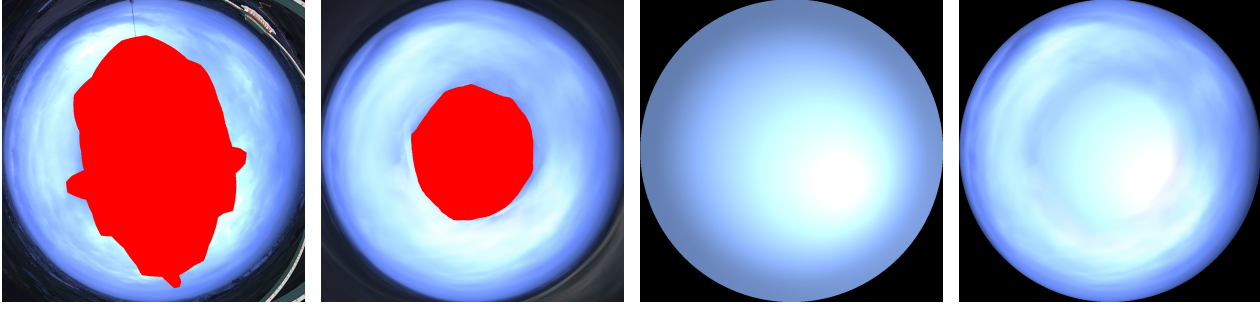
- (1) 地上領域に存在する死角領域の補完
- (2) 空領域に存在する死角領域の輝度値推定
- (3) 残存する死角領域の天空光モデルを用いた補完

(1) 地上領域に存在する死角領域の補完

地上付近の死角領域を補完するため、動画像中の補完対象フレームに対する前フレームにおける輝度情報を利用する。1フレームでの飛行船の移動が十分小さく、死角領域による地上への遮蔽が遠景のみに見られると仮定し、死角周辺の限られた領域を探索することにより、2.3で推定されたカメラの姿勢を用いて方向が合わせられたパノラマ画像座標上における前フレームからの移動ベクトルを推定する。前フレームからの死角領域周辺の移動ベクトルを推定するため、Sum of Squared Difference (SSD) を用いて補完対象フレームにおける死角周辺領域と対応する前フレームにおける領域を推定し、死角領域1つにつき領域の移動ベクトルを1つ求める。死角領域内の画素の輝度値には、推定された移動ベクトルが示す前フレームの対応画素の輝度値を用いる。これを初期フレームから繰り返すことにより、図6(b)のように動画像全体における地上領域に存在する死角領域を補完する。

(2) 空領域に存在する死角領域の輝度値推定

空に存在する死角領域の輝度値を推定するために、動画像全体で空領域における光源環境は不変であり輝度値が変化しないと仮定し、動画像の全フレームから1枚の空画像を生成する。この時、動画像中の各フレームにおける輝度値のみを用いて、死角領域の輝度値を推定する。また空は無縁遠に存在するとして、地上に存在する死角領域の補完時のような画像上での位置合わせを行わずに補完を行う。まず、図5(a)に示されるような、動画像の各フレームの北を上方に合わせた等立体角射影画像を生成する。次に、動画像の全フレームを以下のように統合することにより、1枚の空画像を生成する。統合後の空



(a) 各フレームでの空画像（北が上方の等立体角射影画像、塗りつぶされた領域は死角領域を示す。）
 (b) 動画の全フレームを用いて生成される統合後の空画像
 (c) All Sky Model に基づいて生成される空画像
 (d) 補完後の HDR 空画像（表示輝度値レベルは図 5(a) に統一）

図 5 空領域に存在する死角領域の補完

画像における各画素の輝度値 v_{xy} は、第 i フレームの空画像の輝度値 v_{ixy} を用いて以下のように計算される。

$$v_{xy} = \frac{1}{\sum_i \alpha_{ixy}} \sum_i \alpha_{ixy} \cdot v_{ixy}, \quad (5)$$

$$\alpha_{ixy} = \begin{cases} 1 & (v_{ixy} \text{ が死角領域に含まれない}) \\ 0 & (v_{ixy} \text{ が死角領域に含まれる}) \end{cases}$$

統合後の空画像を図 5(b) に示す。本処理では空領域の光源環境が変化しないと仮定しているため、雲の移動等で光源環境が変化した場合、空領域の輝度勾配が実際より小さくなる。よって、光源環境の大きく変化する環境においては、動画全体で輝度値推定を行わず、数十～数百フレーム程度の短い動画ごとに推定を行うといった対処が考えられる。

(3) 残存する死角領域の天空光モデルを用いた補完

動画を通過して全てのフレームで空が映り込まず死角となる部分では $\sum_i \alpha_{ixy} = 0$ となる。図 5(b) に示すように天頂付近では統合後の空画像の輝度値を決定できないため、全ての天候における空全体の輝度を統計的にモデル化した All Sky Model [10] を用いて補完を行う。All Sky Model は、空全体の輝度分布モデルとして多くの天空状態において実際の観測値に近い近似となっている [10]。All Sky Model は空のおおまかな近似モデルであるため、本処理で補完される領域における雲のテクスチャや局所的な輝度値の変化は再現されない。ここで、計算によって求められる輝度値は元画像における輝度値の上限を上回る可能性があるため、出力される画像を High Dynamic Range(HDR) 画像とする。All Sky Model は、天空の輝度分布を天頂輝度と相対輝度分布の積で表すモデルであり、天空状態（天候等）によって変化するパラメータにより全天候に対応している。All Sky Model は以下のように定式化される。

$$La(\gamma_s, \gamma, \zeta) = Lz(\gamma_s) \cdot L(\gamma_s, \gamma, \zeta)$$

$$= Lz(\gamma_s) \cdot \frac{\phi(\gamma) \cdot f(\zeta)}{\phi(\frac{\pi}{2}) \cdot f(\frac{\pi}{2} - \gamma_s)}, \quad (6)$$

$$\phi(\gamma) = 1 + a \cdot \exp\left(\frac{b}{\sin \gamma}\right),$$

$$f(\zeta) = 1 + c \cdot \{\exp(d \cdot \zeta) - \exp(d \cdot \frac{\pi}{2})\} + e \cdot \cos^2 \zeta,$$

$$a = \frac{6.73}{1 + 0.17 \cdot \exp(3.7 \cdot Si)} - 1.16,$$

$$b = \frac{0.566}{\{1 + 30.1 \cdot \exp(-2.8 \cdot Si)\} - 0.88},$$

$$c = 1.77 \cdot (1.22 \cdot Si)^{3.56} \cdot \exp(0.2 \cdot Si) \cdot (2.1 - Si)^{0.8},$$

$$d = \frac{-3.05}{1 + 10.6 \cdot \exp(-3.4 \cdot Si)},$$

$$e = \frac{0.48}{1 + 245 \cdot \exp(-4.13 \cdot Si)},$$

$La(\gamma_s, \gamma, \zeta)$: 天空要素の輝度,

$Lz(\gamma_s)$: 天頂輝度,

$L(\gamma_s, \gamma, \zeta)$: 相対天空輝度分布,

γ_s : 太陽高度 [rad],

γ : 天空要素の高度 [rad],

ζ : 太陽と天空要素の角距離 [rad],

Si : 天空指標.

ここで、天空指標 Si は全天日射量等から算出され、天候によって $0.0 \leq Si \leq 2.0$ の値をとる。 Si が大きいほど晴天に近く、天空の拡散が少ないことを示す。また、太陽高度 γ_s は撮影日時および緯度・経度から算出される。本来、All Sky Model における天空要素の輝度 La および天頂輝度 Lz には $[W/m^2sr]$ で表される物理量が与えられるが、本研究では物理的な輝度は未知である。よって、本研究ではガンマ補正が行われていない画像を用いることで、画像上での RGB それぞれの輝度値を用いて以降の計算を行う。ここで、ガンマ補正が行われていない画像における飽和が起こっていない画素の輝度値は物理的な輝度に比例するため、相対的な輝度分布は実環境と同じとなる。

本研究では多くの場合天頂が遮蔽されるため、天頂輝度 Lz は未知である。 Lz を推定するため、式 (6) を以下のように変形し、 La および L を導出することで Lz を推定する。

$$Lz(\gamma_s, \gamma, \zeta, Si) = \frac{La(\gamma_s, \gamma, \zeta)}{L(\gamma_s, \gamma, \zeta, Si)}. \quad (7)$$

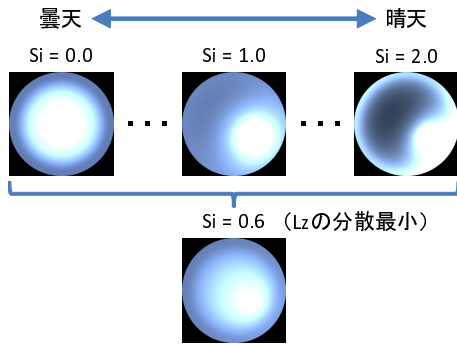


図7 All Sky Model における S_i の推定

ここで、天空指標 S_i は死角領域が存在するために全天日射量等を得ることができず未知であるため、 S_i を変数として扱う。ここで、空画像中における死角領域でない、かつ輝度値の飽和が起こっていない画素をランダムに数百点程度サンプリングし、 k 番目の画素の輝度値を L_{a_k} とする。次に、画素 k における相対輝度 L_k を導出するために、 S_i を決定する必要がある。本研究では、 $S_i = 0.0$ から $S_i = 2.0$ まで 0.1 間隔でサンプリングし、それぞれ独立に $L_k(\gamma_s, \gamma, \zeta, S_i)$ を求め、さらに式 (7) を用いて $L_{z_k}(\gamma_s, \gamma, \zeta, S_i)$ を計算する。図7に示すように、 $L_{z_k}(\gamma_s, \gamma, \zeta, S_i)$ の分散が最も小さくなるように S_i を決定する。決定された S_i を用い、 $L_{z_k}(\gamma_s, \gamma, \zeta, S_i)$ を k について平均することで、 $L_z(\gamma_s, \gamma, \zeta, S_i)$ を推定する。これは、各 S_i に基づいたモデルを用いて空画像を生成し、実際に得られたランダムサンプリングされた点の輝度値との誤差が最小となる S_i を選択することと等価である。

決定された S_i および L_z を用い、死角領域における輝度値を式 (6) から推定する。この際、モデルによって推定された領域と動画像から推定された領域の境界にエッジが発生することがあるため、境界からの距離に応じてアルファ値を低下させることで、両手法によって推定された画素値のブレンディングを行う。

動画像全体で空画像が1枚推定されるため、これを各フレームにおける地上画像と統合することで動画像全体を補完する。地上領域には地上画像の輝度値、空領域には空画像の輝度値を用いることで統合を行う。空画像の補完時と同様に、領域境界付近の画素値についてブレンディングを行う。

仮想物体を重畳する際の背景画像およびレンダリング時の光源として、以上の処理で推定された死角領域による遮蔽のない環境マップを用いる。

2.6 AR シーンのレンダリング

AR シーンのレンダリングには市販のレンダリングソフトウェアである Autodesk 社製 3ds Max および mental image 社製 mentalray を用いる。各フレームのカメラ位置・姿勢により仮想物体の位置合わせを行い、推定された環境マップを用いて IBL を適用した GI により相互反射等を考慮したレンダリングを行い、全方位 AR シーンを生成する。

3. 実 験

3.1 AR シーン生成実験

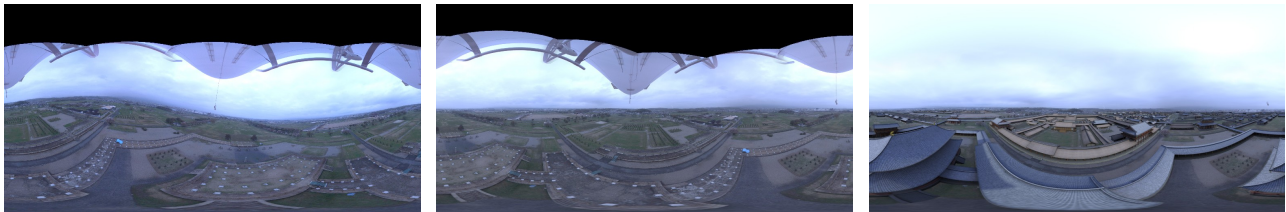
無人飛行船を用いて平城宮跡上空で撮影を行い、仮想物体を合成する実験を行った。実験に用いた動画像はおよそ 1900 フレームであり、約 200 フレームの短い動画像ごとにカメラ位置・姿勢の推定を行った。短い動画像それぞれ、1 フレーム目の姿勢の初期値は姿勢センサの計測値より与えた。飛行船の飛行範囲は南北方向、東西方向ともにおよそ 200m 程度、移動距離はおよそ 600m 程度であった。重畳した仮想物体には、南北 1km、東西 1.3km 程度の平城宮の三次元モデルのうち、建物部分のみを用いた。

合成結果を図8に示す。入力画像を図8(a)に、姿勢を統一した画像を図8(b)に、欠損のない環境マップを用いて生成されたARシーンを図8(c)に示す。All Sky Model における天空指標 S_i を求めた結果、 $S_i = 0.6$ となった。また、文献 [11] を用いて全方位 AR 画像から平面透視投影画像を生成した例を図9に示す。正しく仮想物体の位置合わせが行われ、光源環境についても違和感なく提示されている。

3.2 拡張テレプレゼンスシステムの一般公開実験

平城宮跡上空から撮影された動画像を用いた提案方式に基づくバーチャル歴史体験システムについて、HMD を用いたシステムを開発し、一般公開実験を行った。本実験はおよそ2週間にわたり、平城遷都 1300 年祭の平城宮跡会場内、平城京なりきり体験館で行われ、延べ1000人以上が体験した。生成・蓄積された全方位 AR シーンは、全方位動画像を用いたテレプレゼンスシステム [11] によりユーザの視線方向に追従した平面透視投影画像へ GPU を用いて実時間で変換され、HMD を用いて提示される。また、HMD に内蔵された3軸ジャイロセンサにより HMD の姿勢に応じた動画像を提示することで見回し可能とした。よって、ユーザが実際に見回し動作を行うことにより、その視線方向に対応した平面透視投影画像を得ることができる。システムを用いて提示されるコンテンツには、3.1において撮影され合成されたものを用いた。また、仮想物体が重畳された情景と実環境のみの情景を切り替え可能とし、より直感的に奈良時代当時の風景と実環境を結びつけられるよう配慮を行った。

一般公開実験において各システムを体験したユーザから得られた意見から、実環境の情景と仮想物体を合成して提示することで実際の風景と過去の風景を直感的に結び付けやすいという感想が多く得られた。また、歴史体験システムとして建造物へのアノテーション付与の需要が多く、さらに撮影経路外への視点移動の実現や、仮想物体として動物体を用いることなどを求める意見が多数得られた。仮想物体の位置合わせや陰影表現の統一については概ね良い評価であったが、一部に仮想物体と実物体間のジッタを指摘する意見が見られた。以上より、本



(a) 入力された全方位画像 (b) 姿勢統一後の全方位画像 (c) 生成された全方位 AR 画像

図 8 提案システムにより生成された全方位 AR シーンの例



図 9 全方位 AR シーンから変換された平面透視投影画像の例

研究で目指した幾何学的・光学的整合性等の実現以外に、コンテンツとしての完成度や充実度を高めることも大きな需要の一つであると考えられる。

3.3 拡張テレプレゼンスシステムの定量評価

提案システムの性能を定量的に検証するため、幾何学的整合性を実現するの処理であるカメラ位置・姿勢推定、および光学的整合性を実現するの処理である実物体と仮想物体の陰影関係の統一についての評価を行った。各整合性実現処理に対する評価法および評価結果を以下に述べる。

提案システムのカメラ位置・姿勢推定精度検証のため、再投影誤差を用いて客観的な定量評価を行った。本研究は広域屋外環境を対象としているため、正確な測量によって実空間の3次元座標値の真値を得ることが難しい。そこで、重畳された仮想物体上の画像座標と、対応する実環境の画像座標をそれぞれ手で与えて比較することにより再投影誤差を評価した。本研究では全方位画像を用いており単純に画像座標上での距離を用いて比較することは難しいため、得られた画像座標のカメラを中心とした球面上への射影を求め、各点と球の中心を結ぶ線がなす角を評価に用いた。精度比較の対象として、図12に示すGPSおよび姿勢センサのみを用いてカメラ位置・姿勢を計測して仮想物体を重畳した動画像を用いた。3.1で生成された動画像のうち最初の400フレームを10フレーム毎に、計40フレーム分の再投影誤差を比較した。各フレーム14組の対応を与えた。結果を図10に示す。提案システムの再投影誤差の平均は0.23度、センサのみで仮想物体を重畳した場合の再投影誤差の平均は5.11度であった。有意差が認められ ($p < 0.01$, t検定), 本研究におけるカメラ位置・姿勢推定処理の有効性が示された。

提案システムのカメラ位置の推定精度を別の観点から検証するため、飛行船に搭載されたGPS測位値から計算されたカメラ位置と、SfMを用いて推定されたカメ

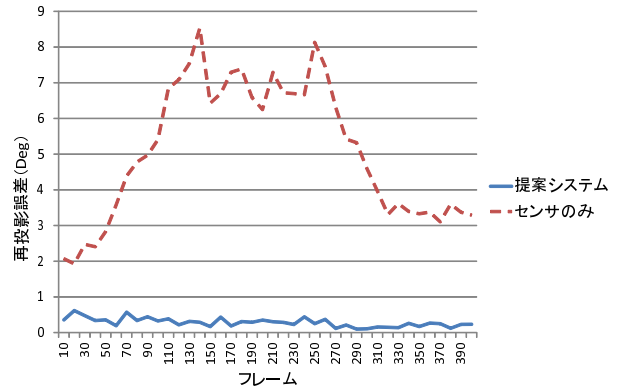


図 10 提案システムにより生成された動画像および比較動画像のカメラ位置・姿勢に基づく再投影誤差

ラ位置のずれを求めた。動画像の前半900フレームにおいて、水平方向の位置ずれ (RMS) が516[mm]、垂直方向の位置ずれ (RMS) が50[mm]となった。動画像前半900フレームの水平、垂直方向のずれを図11に示す。飛行船は急激に高度を変えることはなく、移動方向は主に水平方向であるため垂直方向の位置ずれが比較的小さくなり、移動量の大きい水平方向の位置ずれは比較的大きくなると考えられる。カメラ位置の真値は未知であるが、表1に示したようにGPS受信機の測位精度 (水平方向RMS) はおよそ404[mm]であることから、提案システムで推定された水平方向のカメラ位置精度はおよそ1m程度であると考えられる。カメラの姿勢が正しく推定されていると仮定すると、画像上においては、対象物体までの距離が100mである場合におよそ0.57度、500mの場合に0.11度の誤差となる。これは、上述の再投影誤差とほぼ一致する。

光学的整合性については客観評価を行うことが難しいため、アンケート調査による主観的な定量評価を行うことにより有効性を検証した。撮影日時および撮影地点から得られた太陽の方角からの平行光を光源に用いてレンダリングを行った動画像Aと、提案システムによって生成された動画像Bについて比較を行った。各動画像は、3.2におけるHMDを用いたシステムによりユーザーに提示した。各動画像の見直し動作の後、ユーザーに各動画像の建造物 (仮想物体) と背景 (実物体) 間の陰影関係についての違和感を5段階 (1: 違和感が大きい, 5: 違和感が小さい) で尋ねた。アンケートは20歳代および30歳代の大学院生10名に対して行われた。アンケートの

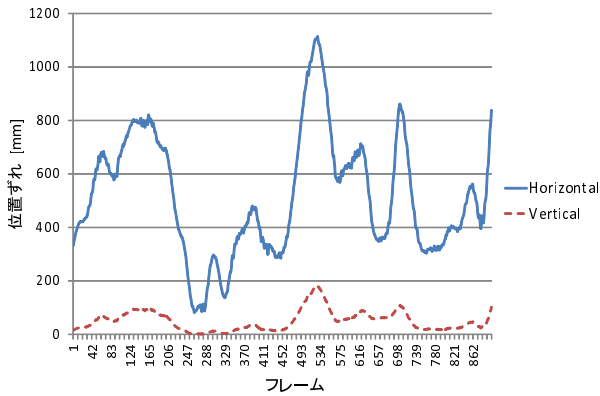
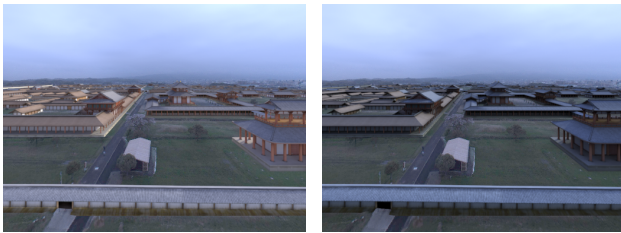


図 11 GPS 測位値から計算されたカメラ位置と SfM を用いて推定されたカメラ位置のずれ



(a) 動画画像 A (b) 動画画像 B

図 12 平行光を光源に用いてレンダリングを行った動画画像 A および提案システムによって生成された動画画像 B から生成した平面透視投影画像 (200 フレーム目)

結果、動画画像 A の評価の平均は 2.1, 提案システムにより生成された動画画像 B の評価の平均は 4.4 となった. 提案システムの評価が高く, 有意差が認められ ($p < 0.01$, Wilcoxon 順位和検定), 提案システムの光学的整合性の実現に関する処理の有効性が示された.

4. まとめと今後の展望

本論文では, 全方位動画画像を用いた見回し可能な蓄積再生型テレプレゼンスに仮想物体を合成する蓄積再生型拡張テレプレゼンスについて, オフラインで高品質に幾何学的・光学的整合性を実現するシステムを提案した. 提案システムでは動画画像と GPS を用いてカメラ位置・姿勢を推定することで幾何学的な位置合わせを行い, 欠損のない環境マップを光源環境として利用してレンダリングを行うことで写実性の高い AR シーンを生成する. 実験では, 平城宮跡上空から撮影された動画画像を用いて AR シーンを生成した. また, 提案方式による蓄積再生型拡張テレプレゼンスの HMD を用いたシステムを構築し, 平城遷都 1300 年祭の平城宮跡会場内, 平城京なりきり体験館において一般公開実験を行い, 延べ 1000 人以上が提案システムを体験した. 一般公開実験における知見から, バーチャル歴史観光システムのコンテンツとしての完成度向上の必要性が挙げられた. 最後に提案システムの幾何学的・光学的整合性の実現に関する処理の性能についての定量評価を行い, 幾何学的整合性の実現に関する処理であるカメラ位置・姿勢推定, および光学的整合性の実現に関する処理である陰影表現の統一について, 提案手法の有効性が示された.

今後は, 提案システムを実時間拡張テレプレゼンスに応用していく予定である. そのためには, 飛行船から地上への全方位動画画像の実時間伝送, 実時間でカメラ位置・姿勢推定, 光源環境推定を行う必要がある. 蓄積再生型システムにおいては, 実空間の幾何形状を考慮することにより, 実物体による仮想物体のオクルージョンの再現を行うことが考えられる. また, 自由視点画像生成を応用した自由な視点移動の実現についても検討を行う.

謝辞 本研究の一部は, 文部科学省特別教育研究経費「アンビエント環境知能研究創出事業」, 科学研究費補助金 (基盤研究 (A), No.23240024) による. 本研究で用いた三次元モデルは凸版印刷株式会社, 社団法人平城遷都 1300 年記念事業協会の提供による.

文 献

- [1] J. Steuer: "Defining virtual reality: Dimensions determining telepresence", *J. of Communication*, **42**, 4, pp. 73-93 (1992).
- [2] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre: "Recent advances in augmented reality", *IEEE Computer Graphics and Applications*, **21**, 6, pp. 34-47 (2001).
- [3] S. Lawson, J. Pretlove, A. Weeler and G. Parker: "Augmented reality as a tool to aid the telerobotic exploration and characterization of remote environments", *Presence: Teleoperators and Virtual Environments*, **2**, 4, pp. 352-567 (2002).
- [4] H. Kim, J. Kim and S. Park: "A bird's-eye view system using augmented reality", *Proc. 32nd Annual Simulation Symposium*, pp. 126-132 (1999).
- [5] F. Okura, M. Kanbara and N. Yokoya: "Augmented telepresence using autopilot airship and omnidirectional camera", *Proc. Ninth IEEE Int. Sympo. on Mixed and Augmented Reality (ISMAR2010)*, pp. 259-260 (2010).
- [6] 島村, 山澤, 竹村, 横矢: "全周パノラマステレオ画像と CG モデルの合成による複合現実環境の構築", *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*, **42**, SIG6(CVIM2), pp. 44-53 (2001).
- [7] P. Ghadirian and I. D. Bishop: "Integration of augmented reality and GIS: A new approach to realistic landscape visualisation", *Landscape and Urban Planning*, **86**, pp. 226-232 (2008).
- [8] Y. Yokochi, S. Ikeda, T. Sato and N. Yokoya: "Extrinsic camera parameter estimation based-on feature tracking and GPS data", *Proc. Seventh Asian Conf. on Computer Vision (ACCV2006)*, Vol. 1, pp. 369-378 (2006).
- [9] S. Ikeda, T. Sato and N. Yokoya: "High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system", *Proc. 2003 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System (MFI2003)*, pp. 155-160 (2003).
- [10] N. Igawa, Y. Koga, T. Matsuzawa and H. Nakamura: "Models of sky radiance distribution and sky luminance distribution", *Solar Energy*, **77**, pp. 137-157 (2004).
- [11] 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).