

NAIST-IS-MT0951019

修士論文

無人飛行船から撮影された全方位動画像を用いた
蓄積再生型拡張テレプレゼンスシステム

大倉 史生

2011年2月3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。

大倉 史生

審査委員：

横矢 直和 教授 (主指導教員)

小笠原 司 教授 (副指導教員)

山澤 一誠 准教授 (副指導教員)

神原 誠之 准教授 (副指導教員)

無人飛行船から撮影された全方位動画像を用いた 蓄積再生型拡張テレプレゼンスシステム*

大倉 史生

内容梗概

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に関する研究が盛んに行われている。さらに、その場所に関連する情報を提示するために、実環境の情景に仮想物体を重畳する拡張現実感 (Augmented Reality: AR) とテレプレゼンスを組み合わせる拡張テレプレゼンスに関する研究が行われている。本論文ではバーチャル歴史観光や都市計画等、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を扱う。本研究では地上での撮影に比べ広範囲を見渡すことのできる空撮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を提案する。本研究で用いる空撮動画像は、低空・低速での撮影に適した無人飛行船を用いて撮影する。本論文では蓄積再生型拡張テレプレゼンスを実現するための技術課題として、(1) 幾何学的整合性の実現、(2) 光学的整合性の実現、(3) テレプレゼンスにおける没入感の向上について述べる。(1) は動画像と GPS を用いてカメラの位置・姿勢を推定することにより現実世界と仮想世界の座標系の位置合わせを行うことによって実現され、(2) は実環境を撮影した全方位動画像を光源環境マップとして利用したレンダリングにより実物体と仮想物体の照明条件を統一することによって解決する。また、(3) については全方位カメラを用いた空撮による自由な見回しを実現し、飛行船の映り込み等に起因する死角領域の消去・補完を行い実環境の遮蔽問題を解決し、さらに全方位動画像の姿勢を統一することにより飛行船の姿勢

*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-MT0951019, 2011年2月3日.

変化によるユーザの意図しない視線方向の変更を抑制する。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 シーンの生成実験とその結果、および平城宮跡で行われた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る。

キーワード

拡張テレプレゼンス, 拡張現実感, テレプレゼンス, 無人飛行船, 幾何学的整合性, 光学的整合性

Augmented Telepresence Using Recorded Aerial Omnidirectional Videos Captured from Unmanned Airship*

Fumio Okura

Abstract

This study is concerned with an augmented telepresence system using aerial images in a large-scale environment. Augmented telepresence provides a user with not only a view of remote site but also related information by augmented reality (AR) technique. An augmented telepresence system using recorded videos can be applied for virtual sightseeing and city planning. This paper proposes an augmented telepresence system which generates a large-scale omnidirectional AR scene in an outdoor environment with a recorded omnidirectional video captured from an unmanned airship. The following technical problems for augmented telepresence with a recorded video are resolved in this research: (1) geometric registration, (2) photometric registration, (3) increasing immersiveness of telepresence. Problem (1) is resolved by extrinsic camera parameter estimation using a video sequence and GPS data, and an environment map of a real scene is used to render virtual objects to overcome the problem (2). As for the problem (3), look-around interaction is enabled by omnidirectional imaging, airship in the omnidirectional image is inpainted to resolve occlusion problem of environment map, and omnidirectional images are aligned to prevent undesired changes of direction due to changes of direction of the airship. This article also shows experimental results

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT0951019, February 3, 2011.

of generating augmented telepresence contents from an omnidirectional video sequence captured on a historic site. Finally, a prototype system for demonstration using the proposed method is described.

Keywords:

Augmented telepresence, Augmented reality, Telepresence, Unmanned airship, Geometric registration, Photometric registration

目次

1. はじめに	1
2. 関連研究と本研究の位置づけ	4
2.1 拡張テレプレゼンスに関する従来研究	4
2.1.1 実時間拡張テレプレゼンス	5
2.1.2 蓄積再生型拡張テレプレゼンス	6
2.2 拡張テレプレゼンスの技術課題に対する従来研究	8
2.2.1 幾何学的整合性の実現	8
2.2.2 光学的整合性の実現	11
2.2.3 蓄積再生型テレプレゼンスにおける没入感向上	14
2.3 本研究の位置づけと解決すべき技術課題	17
3. 飛行船からの空撮全方位動画像を用いた拡張テレプレゼンス	20
3.1 提案手法の概要	20
3.2 全方位動画像の空撮	21
3.3 カメラの位置・姿勢推定	24
3.4 全方位動画像の姿勢統一およびカメラの姿勢補正	27
3.5 死角のない全方位動画像の生成	28
3.5.1 地上領域に存在する死角領域の補完	29
3.5.2 空領域に存在する死角領域の輝度値推定	30
3.5.3 動画像を通して空が映り込まない領域の天空光モデルを用いた補完	31
3.6 AR シーンのレンダリング	34
4. 実験：拡張テレプレゼンスシステムによる AR シーンの生成と提示	36
4.1 提案システムによる AR シーン生成・提示実験	36
4.1.1 全方位動画像の空撮環境および重畳した仮想物体	36
4.1.2 AR シーン生成の実験結果	36
4.2 提案システムを用いた一般公開実験	38

4.3 拡張テレプレゼンスシステムの評価実験	46
5. まとめ	48
謝辞	50
参考文献	51
付録	60
A. 四元数および球面線形補間	60
A.1 四元数を用いた回転表現	60
A.2 球面線形補間	61

目 次

1	Reality-Virtuality (RV) Continuum [5]	5
2	移動体別の空撮高度	6
3	空撮動画像を用いた実時間拡張テレプレゼンスシステム [11]	7
4	自由視点画像生成を応用した蓄積再生型拡張テレプレゼンス [13]	7
5	Google Street View	16
6	提案システムの処理の流れ	21
7	本研究で使用した無人飛行船と搭載機器	23
8	全方位カメラで得られる空撮全方位画像	24
9	動画像と GPS を併用したカメラ位置・姿勢推定 [40] における座標系, 記号および誤差の定義	25
10	全方位画像の姿勢統一	27
11	地上領域に存在する死角領域の補完	29
12	空領域に存在する死角領域の補完	32
13	All Sky Model における S_i の推定	34
14	全方位画像全体の死角領域補完	35
15	平城宮跡上空からの撮影時における飛行船の軌跡	37
16	平城宮の三次元モデル (凸版印刷株式会社製作)	38
17	入力された全方位画像の例	39
18	姿勢統一後の全方位画像の例	40
19	死角領域補完後の全方位画像の例	41
20	提案システムにより生成された全方位 AR シーンの例	42
21	全方位 AR シーンから変換された平面透視投影画像	43
22	一般公開実験に用いた拡張テレプレゼンスシステム	44
23	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	47

表 目 次

1	他航空機と飛行船の比較	5
2	幾何学的整合性実現に関する従来研究	11
3	光源環境取得に関する従来研究	13
4	拡張テレプレゼンスに関する従来研究と本研究の位置付け	17
5	飛行船の仕様	21
6	搭載機器の仕様	22
7	一般公開実験に用いた機器構成	45

1. はじめに

近年、ユーザに遠隔地の情景を没入感高く提示するテレプレゼンス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は、バーチャル観光旅行に代表されるエンタテインメント等、ユーザがオフラインで遠隔地の情景を体感する用途を目的として研究が行われている。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を付加するために、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) 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。本研究ではこれを拡張テレプレゼンス (Augmented Telepresence) と呼ぶ。遠隔地で撮影された情景に実時間で仮想物体を重畳する研究は、ロボット等の遠隔操縦支援や作業支援に用いられる。これに対して、蓄積再生型拡張テレプレゼンスは、建造物の重畳や GIS 情報の可視化を目的として研究が行われている。

本研究では、鳥瞰視点で広範囲の情景を取得するため、飛行船に搭載された全方位カメラを用いて空撮された動画像に建造物等の仮想物体を重畳する蓄積再生型拡張テレプレゼンスシステムを構築する。蓄積再生型拡張テレプレゼンスシステムにより高品質で没入感の高い情景をユーザに提示することで、バーチャル観光旅行等のエンタテインメントや景観シミュレーション・都市計画等への応用が可能である。

飛行船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、以下の技術課題が挙げられる。

1. 幾何学的整合性の実現
2. 光学的整合性の実現
3. 蓄積再生型テレプレゼンスにおける没入感向上

幾何学的整合性は現実環境と仮想環境の座標系位置合わせ問題であり、ビデオスルー拡張現実感の場合、カメラ位置・姿勢を推定することにより実現され

る。従来、広域屋外環境におけるカメラ位置・姿勢推定では GPS と姿勢センサ等によるセンサベースの手法、動画像のみを用いた手法、また、それらを併用したハイブリッド手法について研究が行われている。動画像によるカメラ位置・姿勢推定は仮想物体を重畳する動画像そのものから推定が行われるため仮想物体を重畳する用途に適しているが、事前知識なしでは絶対座標系との関係が未知であるため、スケール情報が得られない。そこで、本研究では GPS と動画像を併用し絶対的なスケール情報付きのカメラ位置・姿勢を推定する手法を用いる。

光学的整合性を実現するための課題の一つとして実物体と仮想物体の陰影表現の統一が挙げられる。従来、拡張現実感等において実シーンに仮想物体を重畳する際の陰影表現に関する研究が行われている。本研究は蓄積再生型システムを想定しているため、実環境の光源環境を用いたグローバルイルミネーションを表現可能であり、比較的計算コストの高い Image-Based-Lighting (IBL) を用いることによりオフラインで高品質なレンダリングを行う。実環境の光源環境を取得するため、本研究では飛行船に全方位カメラを取り付けて撮影を行い、さらに全方位動画像に映り込んだ飛行船および全方位カメラで撮影できない領域からなる死角領域を消去・補完することにより生成された死角のない全方位動画像を光源環境マップとして用いる。

蓄積再生型テレプレゼンスにおける没入感を向上するための課題の一つとして、ユーザの視線方向および視点位置の自由な変更を実現することが挙げられる。これを実現する従来研究として、全方位画像を用いた自由な視線方向変更の実現、自由視点画像の生成による自由な視点移動についての研究が行われている。本研究では、飛行船に取り付けられた全方位カメラを用いて空撮を行うことで視線方向の変更を実現し、さらに飛行船の映り込み等の死角領域を消去・補完する。また、飛行船の姿勢変化によるユーザの意図しない視線方向の変更を抑制するため、全方位動画像の姿勢統一を行う。

本論文では、2章で蓄積再生型拡張テレプレゼンスの技術課題に対する従来の取り組みおよび拡張テレプレゼンスに関する従来研究を概観し、本研究の位置づけと方針を述べる。次に3章で飛行船と全方位カメラを用いた空撮システム、カメラ位置・姿勢の推定による幾何学的整合性の実現、死角領域補完、光学的整合

性の実現法について詳述する．4章では平城宮跡上空から撮影された動画像を用いたARシーンの生成実験とその結果，および平城宮跡で行われた提案システムを用いた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る．最後に5章でまとめおよび今後の展望を述べる．

2. 関連研究と本研究の位置づけ

本章では、拡張テレプレゼンスに関する従来研究、および拡張テレプレゼンスの技術課題に対する従来の取り組みを概観し、本研究の位置づけを述べる。

2.1 拡張テレプレゼンスに関する従来研究

遠隔地の情景を提示するテレプレゼンス [1, 2] と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ビデオシースルー拡張現実感 (Augmented Reality: AR) [3, 4] 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おり、本研究ではこれを拡張テレプレゼンス (Augmented Telepresence) と呼ぶ。テレプレゼンスは現実環境の情景を取り込み仮想環境として扱うため、拡張現実感における仮想物体の重畳が容易に実現できる。蓄積再生型テレプレゼンスは、図 1 に示す Reality-Virtuality (RV) Continuum [5] において仮想化現実 (Augmented Virtuality: AV) に分類される。拡張テレプレゼンスは AR および AV の要素を併せ持つため、図 1 のように RV Continuum において連続体として定義される。特に、実時間拡張テレプレゼンスは現実環境を明示的に仮想化しないため AR にあたる要素が強く、蓄積再生型拡張テレプレゼンスは AV に含まれると考えられる。

本研究では鳥瞰視点で広範囲の情景をユーザに提示可能な空撮動画像の利用に着目する。空撮画像を用いた拡張テレプレゼンスシステムでは、地上の詳細な情報をユーザに提示するために高解像度での空撮が要求される。通常、空撮に多く用いられる有人飛行機の撮影高度は図 2 に示すようにおよそ 1km 以上であることが多いのに対して、無人飛行船や無人ヘリコプタ等の無人航空機 (Unmanned Aerial Vehicle: UAV) の航行高度はおよそ 100m 程度の低空であり、空撮用途として近年注目されている [6, 7]。特に、浮揚のためのエネルギーを必要としない Lighter-Than-Air UAV (LTAUAV) である飛行船は、表 1 における比較でも示されるように、飛行機やヘリコプタより長時間の航行が可能で、折りたたむことで容易に持ち運び可能であるため、空撮への利用が多くなっている [8-10]。そこで本研究では、空撮画像を用いた拡張テレプレゼンスシステムに最適である無人飛行

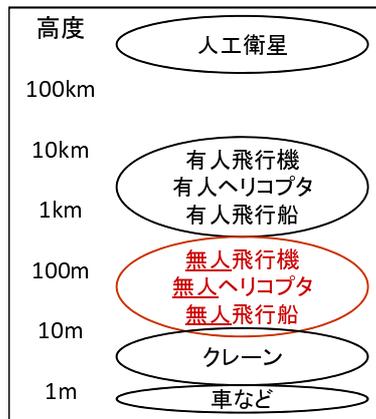
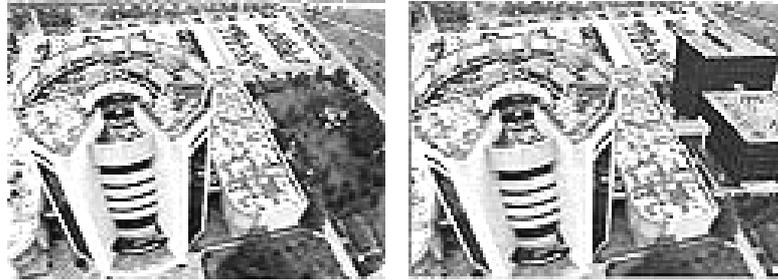


図 2 移動体別の空撮高度

らによる研究 [11] では，無人ラジコンヘリコプタに取り付けられた単眼カメラで図 3(a) のような実環境の情景を取得し，センサで計測されたヘリコプタの位置・姿勢を用いて建造物の 3 次元モデルを図 3(b) のように重畳している．また，Calhoun ら [12] は，遠隔操縦された戦闘機からの偵察等を目的として一般的なディスプレイによる操縦端末を用いて操縦される無人戦闘機から撮影された動画像に実時間で戦闘機の位置や高度情報，GIS による地形情報を重畳するシステムを実現している．

2.1.2 蓄積再生型拡張テレプレゼンス

蓄積再生型拡張テレプレゼンスの研究として文献 [13, 14] が挙げられる．島村ら [13] は，全方位カメラを用いて撮影された画像に仮想物体を重畳するシステムを実現している．視差を設けて 2 枚撮影された全方位画像を用いてテクスチャ付き 3 次元モデルを簡易的に生成することで，図 4 に示すようにユーザの自由な視点移動を実現している．また，Ghadirian ら [14] は，異なる年代の GIS の植生情報から植物のモデルを生成し，あらかじめ撮影された 1 枚のパノラマ画像に重畳することにより，異なる年代の植生を可視化する拡張テレプレゼンスシステムを実現している．



(a) ヘリコプタから取得された実環境の情景 (b) 仮想物体が重畳された情景

図 3 空撮動画像を用いた実時間拡張テレプレゼンスシステム [11]

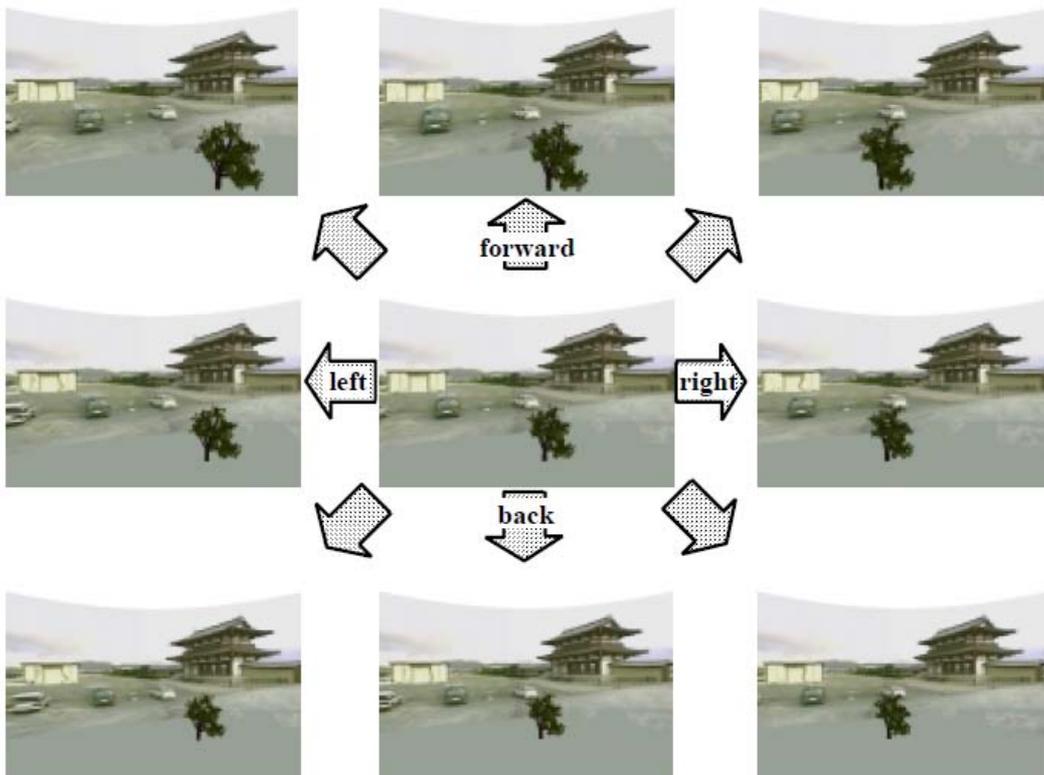


図 4 自由視点画像生成を応用した蓄積再生型拡張テレプレゼンス [13]

2.2 拡張テレプレゼンスの技術課題に対する従来研究

本研究では広域屋外環境において飛行船を用いた蓄積再生型拡張テレプレゼンスを実現する。そのための技術課題として以下の3点が挙げられる。

1. 幾何学的整合性の実現
2. 光学的整合性の実現
3. 蓄積再生型テレプレゼンスにおける没入感向上

幾何学的整合性および光学的整合性の実現は、仮想物体を実環境の情景に正しく重畳するために必要であり、また同時に蓄積再生型テレプレゼンスシステムとしてユーザに高い没入感を与える必要がある。前節で紹介した従来の蓄積再生型拡張テレプレゼンスは、これらの技術課題を解決しているとは言い難い。そこで本節では、これらの技術課題に対する従来研究を概説し、本研究における技術課題解決への道筋を示す。

2.2.1 幾何学的整合性の実現

広域屋外環境における幾何学的整合性を実現するためのカメラ位置・姿勢推定に関する従来研究は、GPSや姿勢センサ等によるセンサベースの手法、動画像のみを用いた手法、それらを併用したハイブリッド手法に大別される。

センサによるカメラ位置・姿勢推定

広域屋外環境におけるセンサを用いたカメラの位置・姿勢推定は絶対座標系上における位置を取得可能なGPSと相対位置・姿勢を得ることのできるジャイロセンサや加速度計を組み合わせることによって実現される。文献 [15-17] では、GPSや姿勢センサ等で得られた位置・姿勢をそのままカメラ位置・姿勢の推定値として用いる。しかし、センサのデータ取得レートがフレームレートより低い場合に全てのフレームにおける正確なカメラの位置・姿勢を推定することが難しい。特に、GPSのデータ取得レートは多くの場合1~数Hzであり、全フレームの位置推定には不十分である。これに対して、神原らによる慣性航法装置の利用 [18],

Ohno らによるオドメトリの利用 [19], Kouroggi らによる歩行動作の解析を用いた GPS や RFID から得られる位置情報の補完 [20] によって全てのフレームにおけるカメラの位置・姿勢推定を行う手法が提案されている。しかし、センサのみを用いる手法では、センサの計測誤差やセンサ・カメラ間のキャリブレーション誤差により高精度なカメラの位置・姿勢推定を行うことは難しい。

動画像のみを用いたカメラ位置・姿勢推定

動画像のみを用いる手法では、画像上の特徴を用いてフレーム間、もしくは事前に用意されたテンプレート等との対応付けを行うことによりカメラの位置・姿勢を推定する。これらの手法は撮影された動画像を用い、比較的高精度にカメラの位置・姿勢を推定できる。

動画像からのカメラ位置・姿勢推定は、事前知識を用いない手法と事前知識を用いる手法に大別できる。ただし、ここで用いる事前知識とは、世界座標系上での 3 次元座標が既知である点群や図形等を画像上で識別可能とした基準マーカを指すものとする。

[事前知識を用いない動画像からのカメラ位置・姿勢推定]

事前知識を用いない動画像からのカメラ位置・姿勢推定は Structure-from-Motion (SfM) と呼ばれる。SfM の一手法として、因子分解法を用いた手法の研究が行われている。因子分解法は、一般的に動画像中の全フレームを通じて全ての特徴点が出現していることを前提としている [21-23]。上島ら [24] は、一部のフレームにおいて特徴点の追跡に失敗した場合、射影復元を利用して特徴点を補完する手法を提案している。しかし、多くの特徴点が出現・消滅するような動画像への適用は難しい。これに対して近年、バンドル調整を用いた手法についての研究が盛んに行われている。これらの手法では、推定された特徴点の三次元位置を画像上へ投影した座標と実際に画像上で検出された特徴点の座標の距離である再投影誤差を最小にするカメラ位置・姿勢および特徴点の三次元位置を求める。バンドル調整により全体最適化を行う手法 [25,26] は、実時間での推定は困難であるが、動画像全体のカメラ位置・姿勢を最適化できるため、高精度な推定が可能である。また、狭区間最適化を逐次的に行うことで実時間でバンドル調整を用いたカメラ

位置・姿勢推定を実現する手法が Klein ら [27] によって提案されている。

一般に、動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 は全体最適化を行い誤差を動画像全体に分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 では原理的に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。

[事前知識を用いた動画像からのカメラ位置・姿勢推定]

3次元位置が既知である点群や図形等の事前知識を用いた手法では、世界座標系上の既知の3次元座標と画像上の2次元座標を対応付け、PnP問題 (Perspective n-Point problem) [28, 29] を解くことによりカメラ位置・姿勢を得る。事前知識を用いる手法としてマーカを用いた手法 [30, 31]、3次元CADモデルを用いる手法 [32, 33]、自然特徴点ランドマークを用いる手法 [34, 35] が研究されている。また、SfM と事前知識を併用することで事前準備にかかる人的コストを低減する研究 [36, 37] が行われている。

事前知識を用いた手法では蓄積誤差が発生せず、スケール既知の世界座標系を用いることで容易にスケール情報を得ることができる。しかし、事前知識となるマーカやCADモデル、ランドマークデータベース等の準備 (事前準備) にかかる人的コストが大きく、特に広域屋外環境においてはコストが膨大となる。

動画像とセンサを併用したカメラ位置・姿勢推定

動画像とセンサを併用したカメラ位置・姿勢推定として、SfM とセンサを組み合わせることにより事前知識なしでスケール情報を得ると同時に蓄積誤差を解消する研究が挙げられる。Agrawal ら [38] および Schleicher ら [39] は、動画像から得られるカメラ位置とGPSの測位位置をカルマンフィルタを用いて融合している。これらの手法は実時間でのカメラ位置を推定することが可能であるが、原理上全体最適化を行うことは難しい。これに対して、拡張バンドル調整を用いた手法 [40–42] が提案されている。これらの手法ではSfMで用いられるバンドル調整の誤差関数にGPSの測位位置に関する誤差の項を追加し全体最適化を行うことで精度向上を図っている。また、事前知識を用いて推定する手法とセンサを組み合わせることにより、カメラの回転に対するロバスト性を向上させることを目的

にした研究 [43–46] も行われている。

本節で紹介した幾何学的整合性実現のための従来研究を表2にまとめる。広域屋外環境への適用を考慮した場合、事前知識を用いる手法における事前準備にかかるコストが膨大になるため、事前知識を用いずスケール情報を得ることができる手法が適していると考えられる。また、蓄積再生型テレプレゼンスにおいては実時間性は必ずしも必要ないため、実時間性は無いが精度の高い手法を適用することができる。よって、広域屋外環境でのカメラの位置・姿勢推定には拡張バンドル調整を用いた手法 [40–42] が最適である。

2.2.2 光学的整合性の実現

光学的整合性における技術課題の一つは実物体と仮想物体の陰影表現の統一である。従来、CGにおける一般的な陰影付けに関する研究は多く行われている [47–49]。これに対し、本研究において実物体と仮想物体の陰影表現を統一するためには、(1) 実環境の光源環境を取得し、(2) 陰影のレンダリングを行う必要がある。これらを実現するため、拡張現実感等の分野における陰影表現に関する研究が行われている。以下、陰影表現統一の各ステップについて従来研究における解決法を述べる。

表 2 幾何学的整合性実現に関する従来研究

手法		精度	実時間性	事前準備 [§]	絶対座標
センサのみ [15–20]		×	✓	△ [†]	✓
動画像のみ	事前知識なし [21–27, 50]	✓	△ [†]	✓	×
	事前知識あり [30–37]	✓	△ [†]	×	△ [†]
ハイブリッド	カルマンフィルタ [38, 39]	△ [‡]	✓	✓	✓
	拡張バンドル調整 [40–42]	✓	×	✓	✓
	回転ロバスト性向上 [43–46]	✓	△ [†]	×	△ [†]

† 手法により異なる。‡ 全体最適化を行わないため精度が劣る。

§ 事前知識構築のための準備（‘✓’: 必要なし, ‘×’: 必要。）

(1) 実環境の光源環境取得

実環境の光源環境を取得するための手法として、仮想物体の重畳に用いる単眼カメラから得られた1枚の画像を用いる手法、シーン内に設置された鏡面球を用いる手法、および仮想物体の重畳に用いるカメラとは別の魚眼レンズ付きカメラを用いる手法が挙げられる。

[単眼カメラから得られた1枚の画像からの光源環境推定手法]

実環境に仮想物体を重畳する最初期の研究である Nakamae ら [51] の手法では、太陽の位置およびカメラ位置・姿勢、実環境の大まかな3次元形状を既知とし1枚の単眼画像から実環境の光源を推定している。実環境の光源を太陽からの直接光の強度および環境光成分のみとし、形状が既知である実環境の任意の2点から推定する。しかし、この手法で用いるのは光源環境の荒い近似であり、さらに推定に最低限の情報のみを与えているため、精度の高い光源環境推定は難しい。Sato ら [52] は、1枚の画像中に映りこんだ実物体の影からシーンの3次元形状を既知として光源環境を推定している。

これらの手法は、単眼カメラから得られる情報量が限られており全方位の光源環境を得ることができない。そのため、撮影シーンや撮影機材に工夫を加え、全方位の光源環境を得る手法に関する研究が行われている。

[鏡面球を用いた光源環境取得]

Debevec ら [53] は、撮影されるシーン内に鏡面球を設置し、これに反射した実環境の情景を光源環境としている。また、神原ら [54] は、マーカと鏡面球を組み合わせることで、幾何学的整合性と光学的整合性を同時に解決する手法を提案している。これらの手法は撮影シーンに工夫を加えることで全方位の情報を得る研究である。しかし、常にカメラから得られた画像に鏡面球が映り込むことが条件であり、広域屋外環境への応用は難しい。

[魚眼レンズ付きカメラからの動画像を用いる手法]

Gibson ら [55] および Agusanto ら [56] は、あらかじめ魚眼レンズ付きカメラ

(ライトプローブ : Light probe) で撮影された HDR 画像を環境マップとして用いることで陰影付けを行っている。佐藤ら [57] は、魚眼レンズ付きカメラを上向きに設置することで実環境の光源環境を実時間で取得している。同様の光源環境取得手法は角田ら [58] の研究にも用いられている。

これらの手法は、鏡面球と比較して環境中の情報を多く得ることができる利点があり、さらにカメラから得られた画像に常に鏡面球が映り込む必要がないため広域屋外環境への適用が容易である。ただし、拡張現実感等において仮想物体の重畳に用いるカメラとは別に環境取得用カメラを設置する必要がある。

光源環境取得に関する従来研究を表 3 にまとめる。鏡面球および魚眼レンズ付きカメラを用いる手法は全方位の光源環境を取得可能である。環境取得用カメラを設置する必要があるが、鏡面球を用いた手法と比較すると広域屋外環境への適用が容易である。

(2) 陰影のレンダリング

仮想物体重畳のために用いる陰影のレンダリング手法として、ローカルイルミネーション (LI) を表現する手法、グローバルイルミネーション (GI) [59] を表現可能な手法が挙げられる。LI は局所照明とも呼ばれ、複数回の反射計算を行わない手法である。それに対して GI は大域照明とも呼ばれ、物体間の相互反射を考慮してレンダリングを行う、LI と比較し写実性の高いシーンを表現可能な手法である。

表 3 光源環境取得に関する従来研究

手法	全方位	広域屋外環境へ適用する際の課題
単眼カメラ [51, 52]	×	特になし
鏡面球 [53, 54]	✓	鏡面球の動画像への映り込みが必要
魚眼レンズ付きカメラ [55-58]	✓	光源環境取得用カメラが必要

[ローカルイルミネーションを表現する陰影付け手法]

Haller ら [60] は、高速な影付けを実現するためシャドウボリューム法を、神原ら [61] は、動的シャドウマップを用いて影付けを行っている。Gibson ら [55] は、シャドウマップを複数用いることで複数の光源によってできる影を再現した。Gibson ら [62] シャドウマップ法を拡張し、面光源の見えを推定することにより従来生成できなかったソフトシャドウを生成している。

[グローバルイルミネーションを表現可能な陰影付け手法]

Fournier ら [63] は実環境の大まかな 3 次元モデルを用いて実物体と仮想物体間の Progressive Radiosity [64] による陰影付けを実現している。Debevec ら [53] はレイトレーシングを用いた GI レンダリングを行っている。これらの研究では、光源環境の表現に High Dynamic Range (HDR) の環境マップを用いる。近年では、これらの GI レンダリング手法が RADIANCE [65] 等のソフトウェアとして実装され、応用分野で多く用いられるようになってきている。また、佐藤ら [66] および Nimeroff ら [67] は、静的環境かつユーザが移動しないことを前提に、あらかじめラジオシティ等で基礎画像をレンダリングしておき、ユーザへの提示画像の生成時に合成することでソフトシャドウを高速に生成している。

GI を表現可能な手法の多くは実時間での陰影付けが難しい。これに対して、LI のみを表現する手法は、現在では実時間もしくは実時間に近い速度で実行可能である。しかし、処理の簡易化により精度が犠牲になる。文献 [66,67] のように、あらかじめ基礎画像等をレンダリングしておくことで GI レンダリングの高速化を図る手法が存在するが、ユーザが移動できないなどの制約を設けているため適用できる環境に限られることが問題となる。

2.2.3 蓄積再生型テレプレゼンスにおける没入感向上

蓄積再生型テレプレゼンスは、現実環境においてユーザが得る情報と遠隔地の情景から得る情報が同じものとなり、さらに現実環境におけるユーザの振る舞いが遠隔地で同様に実現されることが理想である。これらを実現するためには、視

覚的なものに限らず様々な課題が存在する。しかし、蓄積再生型テレプレゼンスに関する研究の多くでユーザによる視線方向もしくは視点位置の自由な変更を扱っている。そこで、以下ではこれらを可能にする研究を概説することにより、蓄積再生型テレプレゼンスの研究を概観する。

従来、蓄積再生型テレプレゼンスにおけるユーザの視線方向および視点位置の変更を可能にするため、全方位カメラによって撮影された画像を用いた視線方向変更の実現、自由視点画像生成を用いた撮影地点外への視点位置変更の実現についての研究が行われている。

全方位カメラを用いた視線方向変更の実現

ユーザの視線方向を変更することによる自由な見回しを実現するため、全方位カメラを用いて撮影された動画像を用いた蓄積再生型テレプレゼンスについて研究が行われている。Onoeら [68] は全方位カメラで撮影された全方位動画像を用い、実時間でユーザの視線方向の平面透視投影画像を生成する研究を行っている。これらの研究ではユーザは全方位カメラで撮影された経路上を自由に移動し、見回すことが可能である。さらに、和田ら [69] は歩行装置を用いた拡張テレプレゼンスシステムにおいて複数経路を移動する場合に、経路の交差点付近で提示する画像を適切に選択することによってシームレスに経路を切り替える研究を行っている。全方位カメラを用いた手法では、全方位カメラの設計や撮影環境によって多くの場合、全方位動画像中に撮影者による情景の遮蔽や全方位カメラで撮影できない部分（死角領域）が生じる。これに対してKawaiら [70] は、地上を歩行しながら撮影された動画像において地面を遮蔽する死角領域を修復し、死角領域のない全方位動画像を生成する手法を提案している。近年では図5に示すGoogle社のGoogle Street Viewのように地図アプリケーションの一機能などとして、自由な視線方向の変更が可能な蓄積再生型テレプレゼンスシステムが一般的に用いられるようになっている。

自由視点画像生成を用いた撮影地点外への視点位置変更の実現

撮影地点以外へのユーザの視点移動を可能にするため、自由視点画像生成を用いたテレプレゼンスについての研究が行われている。島村ら [13] は、シーンの



図 5 Google Street View

3次元情報を仮定し全方位カメラで撮影された画像をテクスチャとして貼り付けることで自由視点画像を生成している。同様に3次元モデルを用いた手法としてAvidanら [71]の研究が挙げられる。一方, Satoら [72]は, 全方位カメラで撮影された動画像を用いて実環境の3次元復元 [73]を行い, 得られた3次元形状からユーザに提示される画像の各画素に届く光線を計算することにより自由視点画像を生成している。また, Horiら [74]は, 自由視点画像生成を応用し, 全方位カメラで撮影された動画像から視差が設けられた画像を生成することにより立体視可能なテレプレゼンスシステムを構築している。

視線方向の変更を可能にする手法と比較し, 自由視点画像生成を用いた手法はユーザに高い自由度を与えるため, 高い臨場感を与えることが可能であると考えられる。自由視点画像生成を行う手法の一部 [13, 72, 74]は全方位カメラで得られた動画像を用いた手法であるため, 従来の全方位動画像を用いたテレプレゼンスに容易に適用可能である。

2.3 本研究の位置づけと解決すべき技術課題

本研究では空撮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を扱う。蓄積再生型拡張テレプレゼンスシステムにより高品質で没入感の高い情景をほぼ全自動で生成・提示することで、バーチャル観光旅行等のエンタテインメントや景観シミュレーション・都市計画等への応用が可能である。従来、表4に示すように蓄積再生型拡張テレプレゼンスの研究として島村らによる研究 [13] や Ghadirian らによる研究 [14] のような1ヶ所で撮影されたパノラマ画像を用いた研究が行われている。しかし、動画像を用い、幾何学的・光学的整合性を実現した蓄積再生型拡張テレプレゼンスシステムは実現されておらず、さらに、蓄積再生型テレプレゼンスシステムとしての没入感向上に関する議論が不十分である。そこで、本研究では幾何学的整合性および光学的整合性を実現し、また、没入感の高いテレプレゼンスを実現する。

蓄積再生型拡張テレプレゼンスを実現するための技術課題は、実環境の情景が撮影された動画像への仮想物体の重畳、および蓄積再生型テレプレゼンスシステムを実現するための没入感の向上に大別される。以下に本研究で取り扱う技術課題と解決方法をまとめる。

表 4 拡張テレプレゼンスに関する従来研究と本研究の位置付け

手法		幾何学的 整合性	光学的整合性		没入感向上
			光源取得	陰影付け	
実時間拡張テレプレゼンス		センサ	×	×	(実時間) †
蓄積再生 型拡張テ レプレゼ ンス	島村ら [13]	×	×	×	自由視点
	Ghadirian ら [14]	×	×	×	単眼カメラ
	本研究	拡張バン ドル調整	魚眼 カメラ	GI	全方位 カメラ

† 実時間拡張テレプレゼンスでは、蓄積再生型拡張テレプレゼンスにおける没入感向上に関する考慮はなされていない。

動画像への仮想物体の重畳

- 幾何学的整合性の実現

[課題 1.1]

仮想物体位置合わせのためのカメラ位置・姿勢推定

[アプローチ 1.1]

動画像と GPS を用いたカメラ位置・姿勢推定手法を適用する。本研究のような広域屋外環境においては、事前準備を行わずにスケール情報を得ることができる手法が適していると考えられる。さらに、本研究では実時間でのカメラ位置・姿勢推定は必要ないため、表 2 に示す従来の幾何学的整合性の実現手法の中でも比較的精度の高い拡張バンドル調整を用い GPS と動画像を併用する Yokochi らの手法 [40] を適用する。また、その際の姿勢の初期値として姿勢センサから得られた値を用いる。

- 光学的整合性の実現

[課題 1.2]

実物体と仮想物体間の陰影表現の統一

[アプローチ 1.2]

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。実環境の光源環境を取得する場合、表 3 に示したように魚眼レンズ付きカメラや鏡面球を用いた撮影は全方位の情報を一度に得ることができる。しかし、鏡面球を用いた手法では、動画像中で常に鏡面球が映り込む必要があり、広域屋外環境下での適用は難しい。よって本研究には魚眼レンズ付きカメラを用いる手法が適している。特に本研究では、撮影に全方位カメラを用いることで仮想物体の重畳に用いるカメラと光源環境を取得する際に同一のカメラを用いる。陰影付けには、実時間ではないが高い品質で陰影付けが可能な GI を用いたレンダリングを行う。

蓄積再生型テレプレゼンスの実現

- 蓄積再生型テレプレゼンスにおける没入感向上

[課題 2.1]

広視野提示による没入感向上

[アプローチ 2.1]

飛行船に搭載された全方位カメラで撮影された実環境の全方位動画像を用いてユーザの視線に追従した透視投影動画像を生成することで、自由な見回しを実現する。2.2.3の蓄積再生型テレプレゼンスの没入感向上に関するカテゴリにおいて、本研究は全方位カメラを用いたテレプレゼンスに該当する。本研究の成果は、自由視点および立体視等を用いた研究に容易に適用可能である。

[課題 2.2]

全方位動画像中の死角領域補完

[アプローチ 2.2]

死角領域を消去・補完することで情景の欠損がない全方位動画像を生成する。Kawaiら [70] は、地上での移動撮影によりに得られた全方位動画像における死角領域を周囲の類似領域を用いて修復しているが、本研究では死角領域の多くは空に存在し大きな運動視差が発生しないため、文献 [70] の手法を適用するのは困難である。そのため、動画像の全フレームから死角領域によって遮蔽された情景を推定し、動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。

[課題 2.3]

全方位動画像の姿勢統一

[アプローチ 2.3]

推定されたカメラの位置・姿勢情報を用いて全方位動画像の姿勢を統一する。また、アプローチ 1.1における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、姿勢統一された動画像を用いて姿勢成分の誤差の補正を行う。

3. 飛行船からの空撮全方位動画像を用いた拡張テレプレゼンス

3.1 提案手法の概要

提案システムの処理の流れを図6に示す。図6では、各処理に対応する前章のアプローチを同時に挙げている。以下に、各処理を順に概説する。

(処理1) 空撮動画像は無人飛行船に搭載された全方位カメラにより撮影される。前章で挙げたテレプレゼンスにおける没入感向上のため、見回し可能な映像を撮影する。また、鳥瞰視点での撮影と組み合わせ広範囲の映像を得る。

(処理2) GPS、姿勢センサ、および空撮動画像を用いてカメラの位置・姿勢を推定することにより、仮想物体を実環境の情景に重畳する際に必要な幾何学的整合性を実現する。

(処理3) カメラの姿勢を用いて全方位動画像の姿勢を統一することにより、飛行船の姿勢変化によるユーザの意図しない視線方向の変更を防止する。同時に、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、姿勢統一された動画像を用いて姿勢成分の誤差の補正を行う。

(処理4) 飛行船の全方位動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・補完することで死角のない全方位動画像を生成する。

(処理5) 最後に、死角のない全方位動画像と撮影位置・姿勢を用いて、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い実環境と仮想環境を合成することで、全方位AR動画像を生成する。ここで、全方位画像を光源環境マップとして用いることで実環境と仮想環境の照明条件を統一する。

以下、全方位動画像の空撮、カメラ位置・姿勢の推定手法、全方位動画像の姿勢統一、死角のない全方位動画像の生成およびARシーンのレンダリングの詳細を述べる。

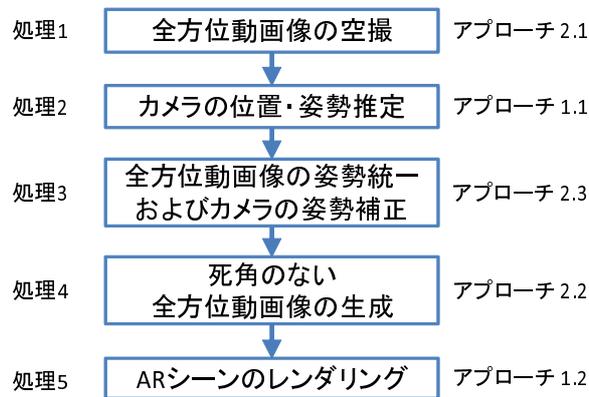


図 6 提案システムの処理の流れ

3.2 全方位動画像の空撮

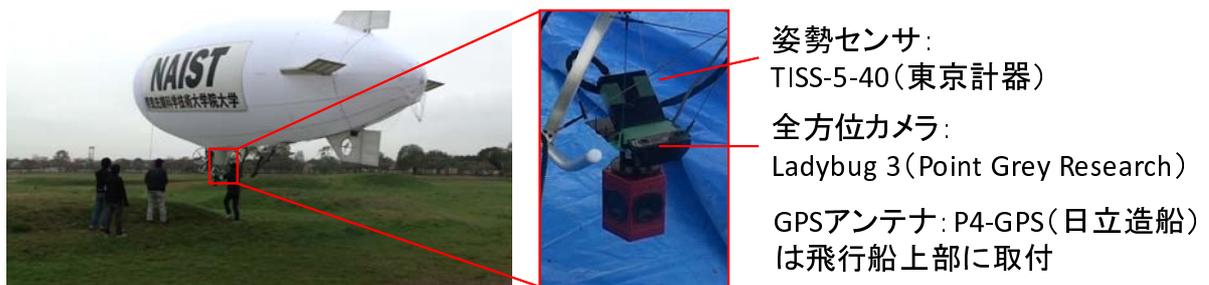
提案システムでは、図 7(a) に示す全長 12m のラジコン飛行船を用いて空撮を行う。飛行船には空撮動画像を取得するための全方位マルチカメラシステム Ladybug3 (Point Grey Research 社製) を下向きに取り付け、離散的な位置情報取得のための GPS として電子基準点網によるネットワーク補正を利用する Differential GPS である P4-GPS (日立造船株式会社製)、姿勢センサとして光ファイバージャイロ TISS-5-40 (東京計器株式会社製) を図 7(b) のように搭載する。全方位カメラによって取得された画像は、全方位カメラにおける各カメラの内部パラメータおよびカメラ間の関係を用いて図 8 に示すようなパノラマ画像に変換される [75]。飛行船の仕様を表 5、各機器の詳細を表 6 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

表 5 飛行船の仕様

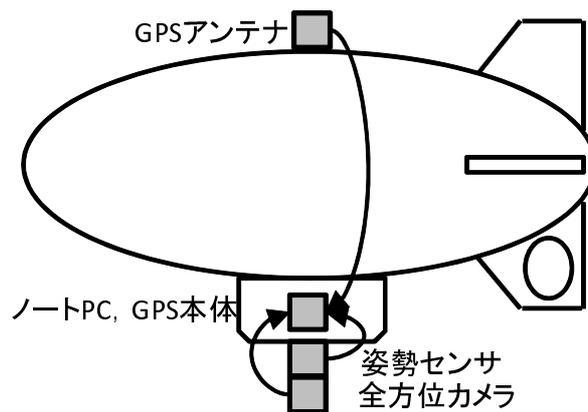
サイズ	全長 12 m, 直径 3.2 m
ペイロード	約 12 kg (気温により変動)
最大速度	50km / h (バッテリー駆動)
連続航行時間	約 60 分

表 6 搭載機器の仕様

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	解像度：5400 × 2700 画素 撮影レート：15 fps 搭載カメラ：6 個 重量：2.4 kg
GPS	日立造船 P4-GPS	計測方式：Differential GPS 計測レート：4Hz 計測誤差：水平 RMS 40.4cm 垂直 RMS 53.5cm 重量：1.1 kg
姿勢センサ	東京計器 TISS-5-40	計測誤差：3°/h 計測レート：500Hz 重量：1.9 kg
PC	DELL XPS M1730	CPU：Intel Core2 Extreme X7900 重量：4.8 kg



(a) 無人飛行船と搭載機器の外観



(b) 機器の搭載位置

図 7 本研究で使用した無人飛行船と搭載機器



図 8 全方位カメラで得られる空撮全方位画像

接続される。

3.3 カメラの位置・姿勢推定

提案システムでは、動画像からカメラ位置・姿勢を推定する Structure-from-Motion (SfM) と GPS の位置情報を併用するハイブリッド手法 [40] を用いてカメラ位置・姿勢を推定する。一般に SfM では蓄積誤差が発生し、さらに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 [40] 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・姿勢推定を行っている。本研究で用いるカメラ位置・姿勢推定手法 [40] 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 9 に示す。一般に SfM 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。

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

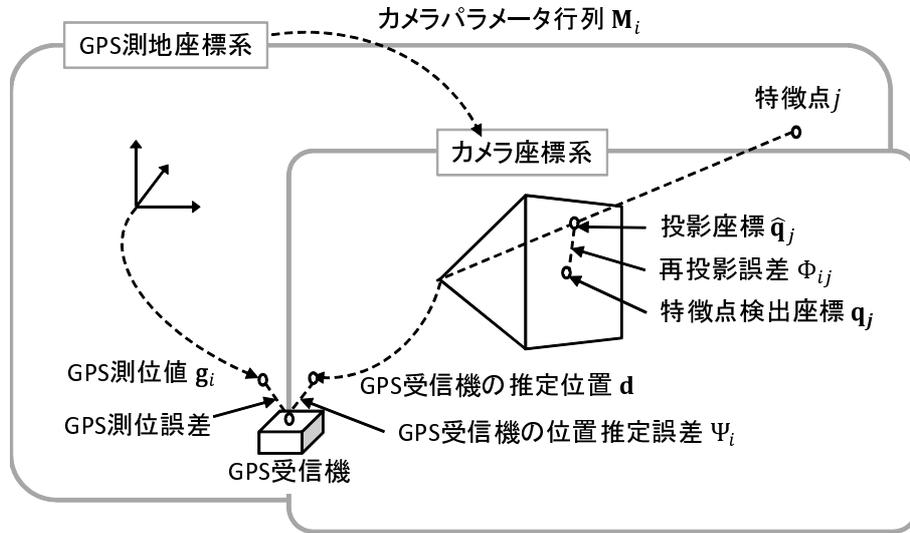


図 9 動画像と GPS を併用したカメラ位置・姿勢推定 [40] における座標系，記号および誤差の定義

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 [40] 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約から 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 [40] 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 [40] 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を切り出して推定を行う。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。なお、四元数についての詳細は付録 A で述べる。

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 [40] 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



図 10 全方位画像の姿勢統一

期値としている。

3.4 全方位動画の姿勢統一およびカメラの姿勢補正

飛行船の姿勢変化によりユーザの注視点が変わることは没入感を損なう原因となりうる。また、本研究では飛行船の映り込みを消去・補完し、ユーザに提示される動画に飛行船が存在しないため、飛行船の姿勢変化は視覚的には画像の揺れとしてユーザに知覚される。そのため、本研究では飛行船の姿勢変化によるユーザの意図しない視線方向の変更を防止するため、3.3 で得られるカメラの位置・姿勢を用いて全方位動画の姿勢統一を行う。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。得られた球面を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を得る。図 8 の全方位動画を姿勢統一したものを図 10 に示す。図 10 は画像の中央が北となるように変換され、地平線が画像上の水平線とほぼ一致していることがわかる。以降のステップでは、姿勢統一後の動画に対して処理を行う。

本研究では遠方のシーンを扱うために 3.3 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してユーザに知覚される。そのため、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. 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。この際、オプティカルフローの算出には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のみを用いる。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。

3.5 死角のない全方位動画像の生成

提案システムで撮影される全方位画像には、飛行船の映り込みや全方位カメラで撮影できない領域が含まれる。本研究ではそれらの領域を死角領域と呼び、死角のない完全な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の実環境の情景への遮蔽に起因す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



(a) 飛行船の映り込みによる地上への遮蔽：
塗りつぶされた領域は死角領域を示す。

(b) 地上領域に存在する死角領域の補完結果：
地上領域に存在する死角領域のみが補完されている。

図 11 地上領域に存在する死角領域の補完

作成し、各フレームの姿勢を用いて3.4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11(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. 動画像を通して空が映り込まない領域の天空光モデルを用いた補完

3.5.1 地上領域に存在する死角領域の補完

地上領域に存在する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。フレーム間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，3.3で推定されたカメラの姿勢を用いて方角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 Differences (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，死角領域に対する領域の移動ベクトルを求める．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 11(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．

3.5.2 空領域に存在する死角領域の輝度値推定

空領域に存在する死角領域の輝度値を推定するために，動画像全体で光源環境は変化せず，空領域の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 12(a) に示すような，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．

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

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

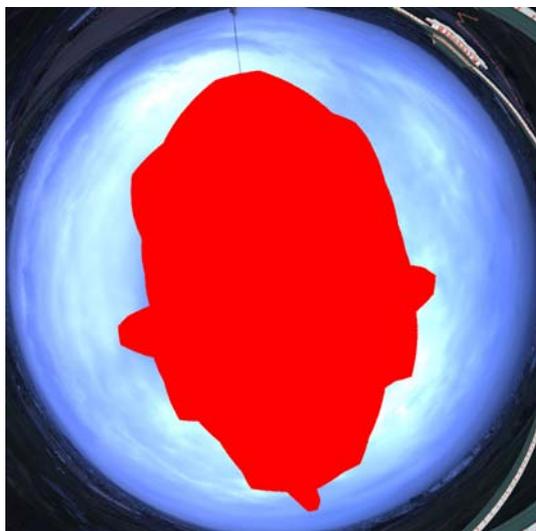
統合後の空画像を図 12(b) に示す．本処理では空領域の光源環境が変化しないと仮定しているため，雲の移動等で光源環境が変化した場合，空領域の輝度勾配が実際より小さくなる．よって，光源環境の大きく変化する環境においては，動画像全体で輝度値推定を行わず，数十～数百フレーム程度の短い動画像ごとに推定するといった対処が考えられる．

3.5.3 動画像を通して空が映り込まない領域の天空光モデルを用いた補完

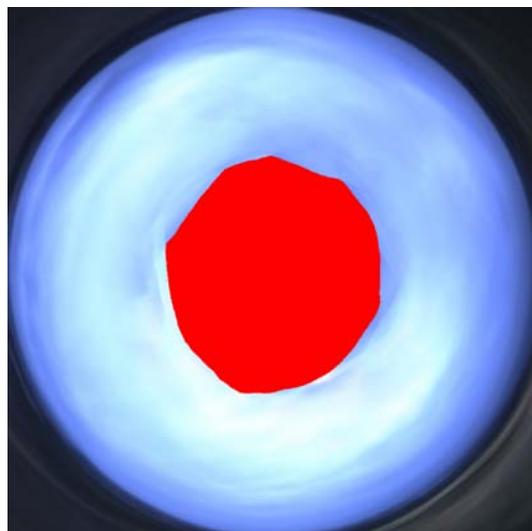
動画像を通して空が映り込まず死角となる部分では $\sum_i \alpha_{ixy} = 0$ となる。図12(b)に示すように天頂付近では統合後の空画像の輝度値を決定できない。本研究では、全ての天候における空全体の輝度を統計的にモデル化した All Sky Model [76] を用いて図12(c)に示す空画像を生成し、図12(d)のように補完を行う。All Sky Model は、空全体の輝度分布モデルとして多くの天空状態において実際の観測値に近い近似となっている [76]。All Sky Model は空のおおまかな近似モデルであるため、本処理で補完される領域における雲のテクスチャや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。

$$\begin{aligned} La(\gamma_s, \gamma, \zeta) &= Lz(\gamma_s) \cdot L(\gamma_s, \gamma, \zeta) \\ &= Lz(\gamma_s) \cdot \frac{\phi(\gamma) \cdot f(\zeta)}{\phi(\frac{\pi}{2}) \cdot f(\frac{\pi}{2} - \gamma_s)}, \end{aligned} \quad (6)$$

$$\begin{aligned} \phi(\gamma) &= 1 + a \cdot \exp\left(\frac{b}{\sin \gamma}\right), \\ f(\zeta) &= 1 + c \cdot \{\exp(d \cdot \zeta) - \exp(d \cdot \frac{\pi}{2})\} + e \cdot \cos^2 \zeta, \\ a &= \frac{6.73}{1 + 0.17 \cdot \exp(3.7 \cdot Si)} - 1.16, \\ b &= \frac{0.566}{\{1 + 30.1 \cdot \exp(-2.8 \cdot Si)\} - 0.88}, \\ c &= 1.77 \cdot (1.22 \cdot Si)^{3.56} \cdot \exp(0.2 \cdot Si) \cdot (2.1 - Si)^{0.8}, \\ d &= \frac{-3.05}{1 + 10.6 \cdot \exp(-3.4 \cdot Si)}, \\ e &= \frac{0.48}{1 + 245 \cdot \exp(-4.13 \cdot Si)}, \\ La(\gamma_s, \gamma, \zeta) &: \text{天空要素の輝度}, \\ Lz(\gamma_s) &: \text{天頂輝度}, \end{aligned}$$



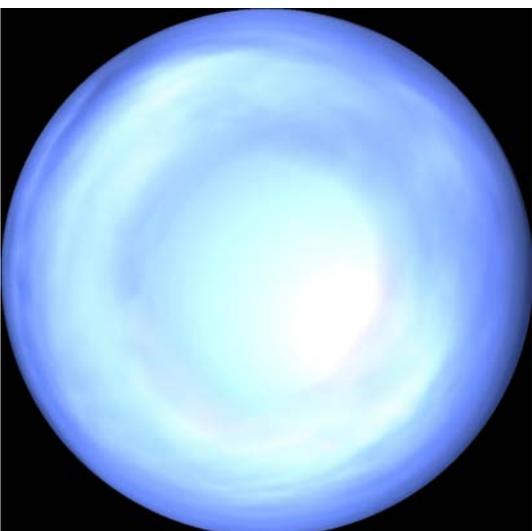
(a) 各フレームでの空画像（北が上方の等立体角射影画像. 塗りつぶされた領域は死角領域を示す.)



(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 12(a) に統一）

図 12 空領域に存在する死角領域の補完

$L(\gamma_s, \gamma, \zeta)$: 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],
 S_i : 天空指標.

ここで、天空指標 S_i は全天日射量等から算出され、天候によって $0.0 \leq S_i \leq 2.0$ の値をとる。 S_i 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 L_a および天頂輝度 L_z には $[\text{W}/\text{m}^2\text{sr}]$ 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、多くの場合に天頂が遮蔽されるため、天頂輝度 L_z は未知である。 L_z を推定するため、式 (6) を以下のように変形し、 L_a および L を導出することで L_z を推定する。

$$L_z(\gamma_s, \gamma, \zeta, S_i) = \frac{L_a(\gamma_s, \gamma, \zeta)}{L(\gamma_s, \gamma, \zeta, S_i)}. \quad (7)$$

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 13 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

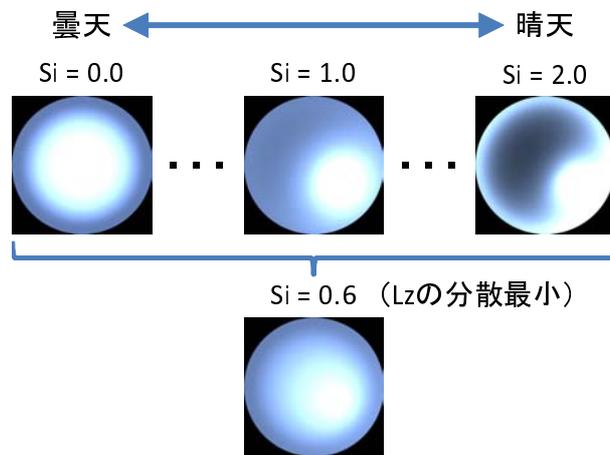


図 13 All Sky Model における Si の推定

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 Si を選択することと等価である。

決定された Si および Lz 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、図 14 のように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。

3.6 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 images 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。カメラ位置・姿勢が表現されている GPS 測地座標系と、



図 14 全方位画像全体の死角領域補完

重畳する仮想物体の座標系を統一する必要があるが、本研究で用いる仮想物体は GPS 測地座標系とスケールおよび座標軸の方向が一致しているため、座標系の原点位置のみを手動で統一している。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、3.5 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。

4. 実験：拡張テレプレゼンスシステムによる AR シーンの生成と提示

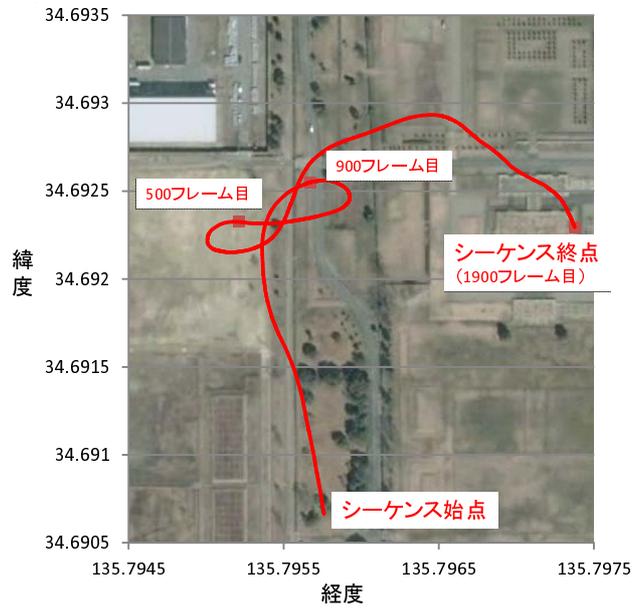
4.1 提案システムによる AR シーン生成・提示実験

4.1.1 全方位動画像の空撮環境および重畳した仮想物体

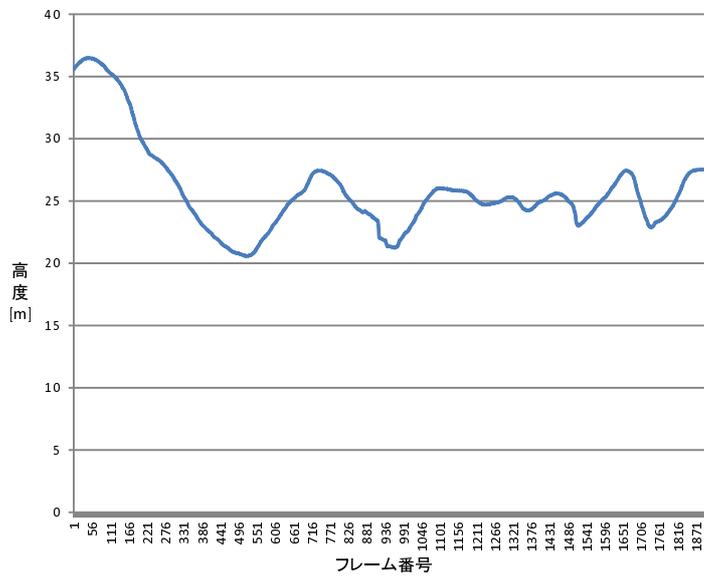
無人飛行船を用いて平城宮跡上空で撮影を行ない、仮想物体を合成する実験を行った。実験に用いた動画像はおよそ 1900 フレームからなり、約 200 フレームの短い動画像ごとにカメラ位置・姿勢の推定を行った。短い動画像それぞれ、1 フレーム目の姿勢の初期値は姿勢センサの計測値より与えた。飛行船に搭載された GPS によって得られた飛行パスを図 15 に示す。飛行船の飛行範囲は南北方向、東西方向ともにおよそ 200m 程度、移動距離はおよそ 600m 程度であった。重畳した仮想物体には、図 16 に示す南北 1km、東西 1.3km 程度の平城宮の三次元モデルのうち、建物部分のみを用いた。図 16 は奈良時代後期の平城宮を復元したモデルであり、凸版印刷株式会社が製作し平城遷都 1300 年祭における平城京 VR シアターで用いられたものである。

4.1.2 AR シーン生成の実験結果

平城宮跡上空の飛行船から撮影された全方位動画像を用いた全方位 AR シーンの生成結果を示す。提案手法により、入力動画像と同様に 1900 フレームの AR シーンを生成した。入力画像例を図 17 に、3.3 で推定されたカメラ位置・姿勢を用いて 3.4 で姿勢統一を行った画像例を図 18 に示す。地平線が画像の水平線と一致していることがわかり、正しく姿勢統一が行われていることが確認できる。また、図 19 に示す 3.5 における死角領域の補完結果より、地上領域および空領域ともに正しく死角領域の補完が行われていることがわかる。この時、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。生成された AR シーンの例を図 20 に示す。また、Onoe らの手法 [68] 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 21 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ことが確認できる。しかし、生成



(a) 飛行船の軌跡（水平方向）：緯度は 0.001° あたり約 111m, 経度は 0.001° あたり約 90m.



(b) 飛行船の軌跡（高度方向）

図 15 平城宮跡上空からの撮影時における飛行船の軌跡



図 16 平城宮の三次元モデル（凸版印刷株式会社製作）

された動画像にはジッタが発生しており，違和感の原因となる可能性がある．姿勢誤差の補正は行ったが位置誤差の補正を行っていないため，位置誤差に起因すると考えられる．また，実環境の幾何形状を考慮していないため，オクルージョン関係等が正しく反映されていない．本実験のように実環境が平面に近い場合には大きな問題にならないが，実環境の形状によっては問題となる．

4.2 提案システムを用いた一般公開実験

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，TV モニタを用いたシステムと 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．公開実験に用いたシステムの外観を図 22 に示す．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 [68] 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 等のディスプレイデバイスを用いて提示される．TV モニタを用いたシステムでは，図 22(a) のように全方位動画像から切り出された平面透視投影動画像を，一般的に販売されている比較的大きな TV モニタに提示し，ジョイスティックを用いて見回しを実現した．HMD を用いたシステムでは，図 22(b) のように



500フレーム目

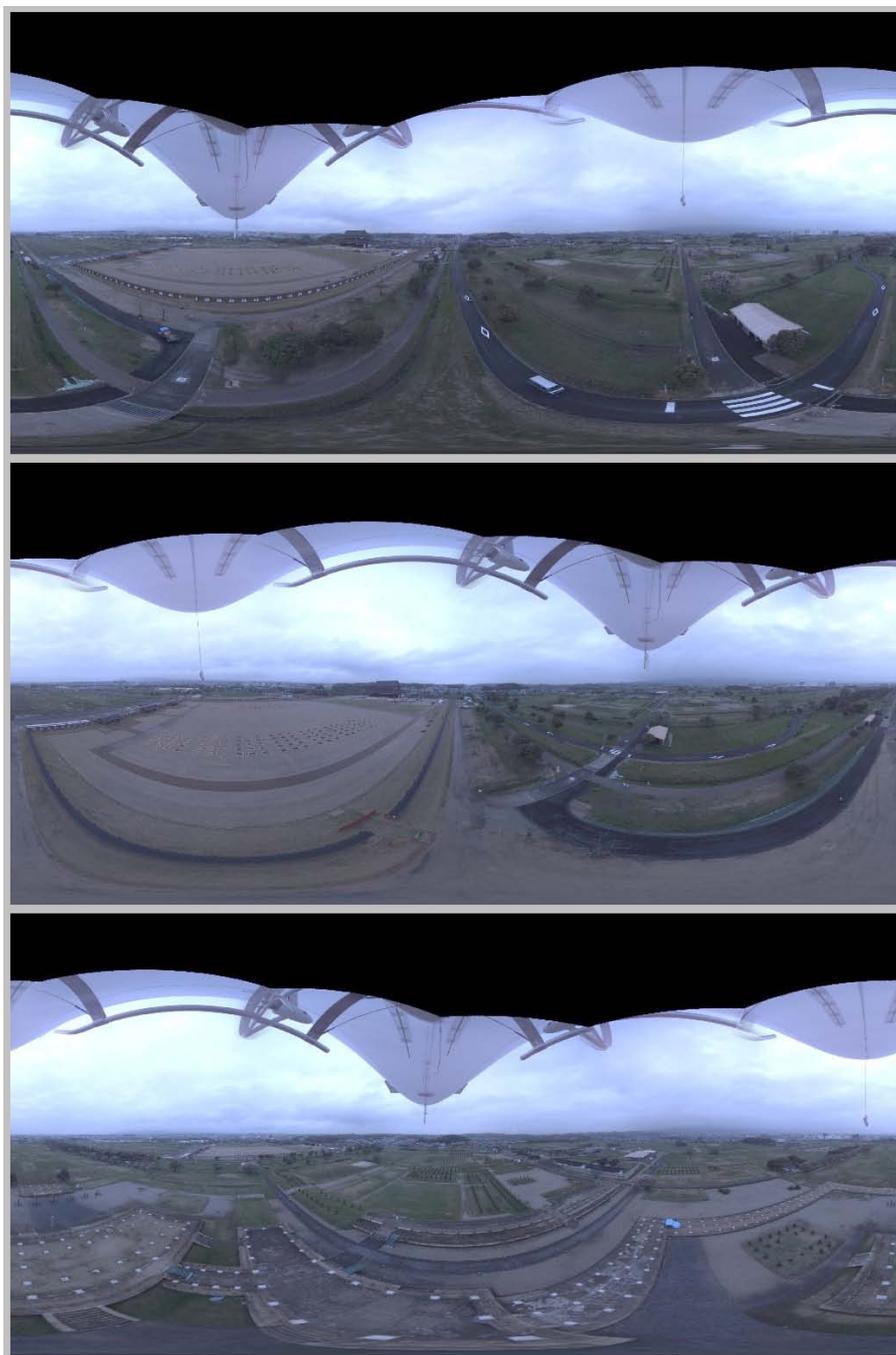


900フレーム目



1900フレーム目

図 17 入力された全方位画像の例



500フレーム目

900フレーム目

1900フレーム目

図 18 姿勢統一後の全方位画像の例



500フレーム目



900フレーム目



1900フレーム目

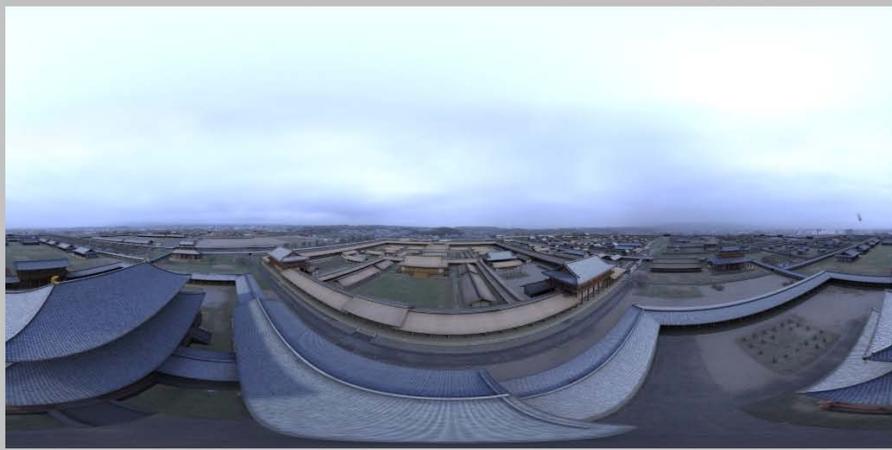
図 19 死角領域補完後の全方位画像の例



500フレーム目



900フレーム目



1900フレーム目

図 20 提案システムにより生成された全方位 AR シーンの例



図 21 全方位 AR シーンから変換された平面透視投影画像



(a) TV モニタを用いたシステム

(b) HMD を用いたシステム

図 22 一般公開実験に用いた拡張テレプレゼンスシステム

生成された平面透視投影動画は HMD に提示され、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を提示することで見回し可能とした。各システムに用いた機器の詳細を表 7 に示す。システムを用いて提示されるコンテンツには、4.1 において撮影され合成されたものを用いた。ただし、一般公開実験時点では、3.4 におけるカメラの姿勢の誤差補正は行っていない。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より、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かった。また、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、3.4 におけるカメラ姿勢の誤差補正が未実装であったため、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。

表 7 一般公開実験に用いた機器構成

	機器	モデル	詳細
HMD を 用いた システム	PC	DELL Vostro 1400	CPU : Intel Core2 Duo T7300 グラフィック : チップセットに内蔵
	HMD	iWear VR920	解像度 : 640×480 視野角 : 32° 3軸ジャイロセンサ内蔵
TV モニタ を用いた システム	PC	Faith MTX2 i7720	CPU : Intel Core i7 X920 グラフィック : GeForce GTX 285M
	TV モニタ	シャープ AQUOS LC-65RX1W	画面サイズ : 65 インチ 解像度 : 1920×1080

4.3 拡張テレプレゼンスシステムの評価実験

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ため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ため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動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、GPSおよび姿勢センサのみで仮想物体を重畳した動画像を用いた。4.1.2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、各組の再投影誤差をフレームごとに平均した結果を図23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。

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、4.2におけるHMDを用いたシステムによりユーザに提示した。各動画像の見直し動作の後、ユーザ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。アンケートの結果、動画像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、

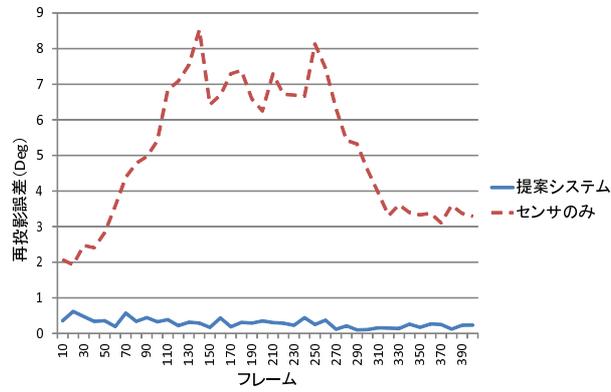


図 23 提案システムにより生成された動画像および比較動画像のカメラ位置・姿勢に基づく再投影誤差

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.

5. まとめ

本論文では、幾何学的・光学的整合性を実現し、さらに没入感の高いテレプレゼンスを実現する蓄積再生型拡張テレプレゼンスシステムを提案した。全方位カメラをGPS、姿勢センサとともに無人飛行船に搭載し、他の移動体と比較して低空・低速での空撮を行った。提案システムでは動画像とGPSを併用し絶対座標系上でのカメラの位置・姿勢を推定する [40] ことで、幾何学的整合性を実現した。同時に、推定された全方位カメラの姿勢情報を用いて全方位動画像の姿勢統一を行い、ユーザの意図しない視線方向の変更を抑制した。また、全方位動画像への飛行船の映り込み等の死角領域が補完された全方位動画像を光源環境マップ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シーンを生成した。死角領域の補完は動画像の全フレームを用いて行われ、動画像を通じて死角となる部分では天空輝度分布モデルである All Sky Model [76] を用いて補完を行った。

実験では、平城宮跡上空から撮影された動画像を用いて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およびディスプレイ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。

本研究は蓄積再生型拡張テレプレゼンスシステムを扱っているが、今後の展望として実時間拡張テレプレゼンスへの発展が考えられる。本研究に基づく実時間システムを構築する場合、各技術課題を解決するアプローチには実時間性のある手法を用いる。特に、カメラ位置・姿勢推定および陰影表現の統一についての検討が必要である。カメラ位置・姿勢推定については、実時間性がありかつ事前準備の必要ないカルマンフィルタを用いた手法や、センサを用いた手法を適用することが考えられる。陰影表現の統一については、拡張現実感において研究が行われている高速な陰影付け手法が適用できる。また、実時間システムでは、撮影地

点からユーザへ実時間で動画像を伝送する必要がある。特に全方位動画像の伝送には広帯域が必要であるが、空撮を行う場合は有線伝送は困難であり、伝送手法に関する検討が必要である。また、実験より位置誤差に起因するジッタが残存しているため、カメラの位置・姿勢推定の改良や補正処理を行うことが考えられる。実空間の幾何形状を考慮することによりオクルージョンの再現等を行うことができる。

謝辞

本研究の全過程を通して、懇切なる御指導、御鞭撻を賜りました視覚情報メディア講座 横矢 直和 教授に心より感謝致します。また、本研究の遂行にあたり、有益な御助言、御鞭撻を頂いたロボティクス講座 小笠原 司 教授に厚く御礼申し上げます。そして、本研究を進めるにあたり、始終暖かい御指導をして頂いた視覚情報メディア講座 山澤 一誠 准教授に深く感謝致します。また、本研究を行うにあたり、多大なる御助言、御鞭撻を賜った環境知能学講座 神原 誠之 准教授に心より感謝致します。さらに、本研究を通じて、的確な御助言、御鞭撻を頂いた視覚情報メディア講座 佐藤 智和 助教に深く御礼申し上げます。特に、神原 誠之 准教授には本研究の着想およびテーマ設定から研究の遂行、論文執筆、発表練習その他公私にわたる様々なご指導をいただきました。また、研究室生活において様々な支援をして頂いた、視覚情報メディア講座秘書 中村 美奈 女史および前秘書である 高橋 美央 女史に厚く御礼申し上げます。最後に、研究のみならず研究室生活全般においてお世話になりました視覚情報メディア講座の諸氏に深く感謝いたします。

参考文献

- [1] M. Minsky. Telepresence. *Omni*, pp. 45–51, 1980.
- [2] S. Moezzi. Special issue on immersive telepresence. *IEEE Multimedia*, Vol. 4, No. 1, pp. 17–56, 1997.
- [3] R. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, pp. 355–385, 1997.
- [4] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, Vol. 21, No. 6, pp. 34–47, 2001.
- [5] P. Milgram. A taxonomy of mixed reality visual displays. *IEICE Trans. on Information Systems*, Vol. E77-D, No. 9, pp. 1321–1329, 1994.
- [6] S. R. Herwitz, L. F. Johnson, S. E. Dunagan, R. G. Higgins, D. V. Sullivan, J. Zheng, B. M. Lobitz, J. G. Leung, B. A. Gallmeyer, M. Aoyagi, R. E. Slye, and J. A. Brass. Imaging from an unmanned aerial vehicle: Agricultural surveillance and decision support. *Computers and Electronics in Agriculture*, Vol. 44, pp. 49–61, 2004.
- [7] M. A. Goodrich, B. S. Morse, D. Gerhardt, and J. L. Cooper. Supporting wilderness search and rescue using a camera-equipped mini UAV. *J. of Field Robotics*, Vol. 25, No. 1, pp. 89–110, 2008.
- [8] E. Paiva, J. Azinheira, J. Ramos, A. Moutinho, and S. Bueno. Project AURORA: Infrastructure and flight control experiments for a robotic airship. *J. of Field Robotics*, Vol. 23, No. 2–3, pp. 201–222, 2006.
- [9] T. Fukao, A. Yuzuriha, T. Suzuki, T. Kanzawa, T. Oshibuchi, K. Osuka, T. Kohno, M. Okuyama, Y. Tomoi, and M. Nakadate. Inverse optimal velocity field control of an outdoor blimp robot. In *Proc. 17th IFAC World Congress*, pp. 4374–4379, 2008.

- [10] E. Hygounenc, I. K. Jung, P. Soueres, and S. Lacroix. The autonomous blimp project of LAAS-CNRS: Achievements in flight control and terrain mapping. *Int. J. of Robotics Research*, Vol. 23, No. 4–5, pp. 473–511, 2004.
- [11] H. Kim, J. Kim, and S. Park. A bird’s-eye view system using augmented reality. In *Proc. 32nd Annual Simulation Symposium*, pp. 126–132, 1999.
- [12] G. L. Calhoun, M. H. Draper, M. F. Abernathy, F. Delgado, and M. Patzek. Synthetic vision system for improving unmanned aerial vehicle operator situation awareness. In *Proc. SPIE Enhanced and Synthetic Vision 2005*, Vol. 5802, pp. 219–230, 2005.
- [13] 島村潤, 山澤一誠, 竹村治雄, 横矢直和. 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*, Vol. 42, No. SIG6(CVIM2), pp. 44–53, 2001.
- [14] P. Ghadirian and I. D. Bishop. Integration of augmented reality and GIS: A new approach to realistic landscape visualisation. *Landscape and Urban Planning*, Vol. 86, pp. 226–232, 2008.
- [15] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal and Ubiquitous Computing*, Vol. 1, No. 4, pp. 208–217, 1997.
- [16] R. Azuma, B. Hoff, H. Neely III, and R. Sarfaty. A motion-stabilized outdoor augmented reality system. In *Proc. IEEE Virtual Reality (VR’99)*, pp. 252–259, 1999.
- [17] W. Piekarski, R. Smith, and B.H. Thomas. Designing backpacks for high fidelity mobile outdoor augmented reality. In *Proc. Third IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2004)*, pp. 280–281, 2004.

- [18] 神原誠之, 横矢直和. RTK-GPS と慣性航法装置を併用したハイブリッドセンサによる屋外型拡張現実感システム. 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 933–938, 2005.
- [19] K. Ohno, T. Tsubouchi, B. Shigematsu, S. Maeyama, and S. Yuta. Differential GPS and odometry-based outdoor navigation of a mobile robot. *Advanced Robotics*, Vol. 18, No. 6, pp. 611–635, 2004.
- [20] M. Kouroggi, N. Sakata, T. Okuma, and T. Kurata. Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/self-contained sensor system. In *Proc. 16th Int. Conf. on Artificial Reality and Telexistence (ICAT2006)*, pp. 1310–1321, 2006.
- [21] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. of Computer Vision*, Vol. 9, No. 2, pp. 137–154, 1992.
- [22] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. Fourth European Conf. on Computer Vision (ECCV'96)*, pp. 709–720, 1996.
- [23] 植芝俊夫, 富田文明. 奥行パラメータの逐次推定による多視点透視投影画像のための因子分解法. 電子情報通信学会論文誌 (D), Vol. J81-D2, No. 8, pp. 1718–1726, 1998.
- [24] 上島重治, 斎藤英雄. 因子分解法のための射影復元を利用した計測行列の補間法. 画像電子学会誌, Vol. 33, No. 4B, pp. 576–585, 2004.
- [25] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment—a modern synthesis. In *Proc. Int. Workshop on Vision Algorithms 1999*, pp. 278–372, 1999.
- [26] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *Int. J. of Computer*

- Vision*, Vol. 59, No. 3, pp. 207–232, 2004.
- [27] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. Sixth IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2007)*, pp. 1–10, 2007.
- [28] R. Horaud, B. Conio, O. Le Boulleux, and B. Lacolle. An analytic solution for the perspective 4-point problem. *Computer Vision, Graphics, and Image Processing*, Vol. 47, No. 1, pp. 33–44, 1989.
- [29] R. Klette, A. Koschan, and K. Schluns. *Computer Vision: Three-dimensional Data from Images*. Springer, 1998.
- [30] 加藤博一, M. Billingham, 浅野浩一, 橘啓八郎. マーカー追跡に基づく拡張現実感システムとそのキャリブレーション. 日本バーチャルリアリティ学会論文誌, Vol. 4, No. 4, pp. 607–616, 1999.
- [31] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski. ARQuake: An outdoor/indoor augmented reality first person application. In *Proc. Sixth IEEE Int. Sympo. on Wearable Computers (ISWC2002)*, pp. 139–146, 2002.
- [32] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 932–946, 2002.
- [33] L. Vacchetti, V. Lepetit, and P. Fua. Combining edge and texture information for real-time accurate 3D camera tracking. In *Proc. Third IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2004)*, pp. 48–57, 2004.
- [34] E. Royer, M. Lhuillier, M. Dhome, and J.M. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *Int. J. of Computer Vision*, Vol. 74, No. 3, pp. 237–260, 2007.

- [35] 武富貴史, 佐藤智和, 横矢直和. 拡張現実感のための優先度情報を付加した自然特徴点ランドマークデータベースを用いた実時間カメラ位置・姿勢推定. 電子情報通信学会論文誌 (D), Vol. 92, pp. 1440–1451, 2009.
- [36] 佐藤智和, 神原誠之, 横矢直和, 竹村治雄. マーカと自然特徴点の追跡による動画像からのカメラ移動パラメータの復元. 電子情報通信学会論文誌 (D-II), Vol. 86, pp. 1431–1440, 2003.
- [37] G. Bleser, H. Wuest, and D. Stricker. Online camera pose estimation in partially known and dynamic scenes. In *Proc. Fifth IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2006)*, pp. 56–65, 2006.
- [38] M. Agrawal and K. Konolige. Real-time localization in outdoor environments using stereo vision and inexpensive GPS. In *Proc. 18th Int. Conf. on Pattern Recognition (ICPR2006)*, Vol. 3, pp. 1063–1068, 2006.
- [39] D. Schleicher, L.M. Bergasa, M. Ocana, R. Barea, and M.E. Lopez. Real-time hierarchical outdoor SLAM based on stereovision and GPS fusion. *IEEE Trans. on Intelligent Transportation Systems*, Vol. 10, No. 3, pp. 440–452, 2009.
- [40] Y. Yokochi, S. Ikeda, T. Sato, and N. Yokoya. Extrinsic camera parameter estimation based-on feature tracking and GPS data. In *Proc. Seventh Asian Conf. on Computer Vision (ACCV2006)*, Vol. 1, pp. 369–378, 2006.
- [41] T. Anai, N. Fukaya, T. Sato, N. Yokoya, and N. Kochi. Exterior orientation method for video image sequences with considering RTK-GPS accuracy. In *Proc. Ninth Int. Conf. on Optical 3-D Measurement Techniques*, Vol. 1, pp. 231–240, 2009.
- [42] H. Kume, T. Taketomi, T. Sato, and N. Yokoya. Extrinsic camera parameter estimation using video images and GPS considering GPS positioning accuracy. In *Proc. 20th IAPR Int. Conf. on Pattern Recognition (ICPR2010)*, pp. 3923–3926, 2010.

- [43] B. Jiang, U. Neumann, and S. You. A robust hybrid tracking system for outdoor augmented reality. In *Proc. IEEE Virtual Reality 2004 (VR2004)*, pp. 3–10, 2004.
- [44] G. Reitmayr and T.W. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *Proc. Sixth IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2007)*, pp. 109–118, 2007.
- [45] G. Klein and T. Drummond. Robust visual tracking for non-instrumented augmented reality. In *Proc. Second IEEE and ACM Int. Sympo. on Mixed and Augmented Reality (ISMAR2003)*, pp. 113–122, 2003.
- [46] 松田幸大, 池田聖, 佐藤智和, 横矢直和. 自然特徴点ランドマークデータベースと姿勢センサに基づく高速な回転にロバストなカメラ位置・姿勢推定. 画像の認識・理解シンポジウム (MIRU2007) 講演論文集, pp. 1558–1563, 2007.
- [47] F. C. Crow. Shadow algorithms for computer graphics. In *Proc. ACM SIGGRAPH '77*, pp. 242–248, 1977.
- [48] A. Woo, P. Poulin, and A. Fournier. A survey of shadow algorithms. *IEEE Computer Graphics and Applications*, Vol. 10, No. 6, pp. 13–32, 1990.
- [49] J. M. Hasenfratz, M. Lapierre, N. Holzschuch, and F. Sillion. A survey of real-time soft shadows algorithms. *Computer Graphics Forum*, Vol. 22, No. 4, pp. 753–774, 2003.
- [50] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, pp. 1052–1067, 2007.
- [51] E. Nakamae, K. Harada, T. Ishizaki, and T. Nishita. A montage method: The overlaying of the computer generated images onto a background photograph. In *Proc. ACM SIGGRAPH '86*, pp. 207–214, 1986.

- [52] I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 3, pp. 290–300, 2003.
- [53] P. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. ACM SIGGRAPH '98*, pp. 189–198, 1998.
- [54] M. Kanbara and N. Yokoya. Geometric and photometric registration for real-time augmented reality. In *Proc. the First International Symposium on Mixed and Augmented Reality (ISMAR2002)*, pp. 279–280, 2002.
- [55] S. Gibson and A. Murta. Interactive rendering with real-world illumination. In *Proc. Eurographics Workshop on Rendering Techniques*, pp. 365–376, 2000.
- [56] K. Agusanto, L. Li, Z. Chuangui, and N.W. Sing. Photorealistic rendering for augmented reality using environment illumination. In *Proc. Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pp. 208–216, 2003.
- [57] I. Sato, Y. Sato, and K. Ikeuchi. Acquiring a radiance distribution to superimpose virtual objects onto a real scene. *IEEE Trans. on Visualization and Computer Graphics*, Vol. 5, pp. 1077–2626, 2002.
- [58] 角田哲也, 大石岳史, 池内克史. 影付け平面を用いた複合現実感における高速陰影表現手法. *映像情報メディア学会誌*, Vol. 62, No. 5, pp. 788–795, 2008.
- [59] H. W. Jensen. Global illumination using photon maps. In *Proc. Eurographics Workshop on Rendering*, pp. 21–30, 1996.
- [60] M. Haller, S. Drab, and W. Hartmann. A real-time shadow approach for an augmented reality application using shadow volumes. In *Proc. 10th ACM*

Sympo. on Virtual Reality Software and Technology (VRST2003), pp. 56–65, 2003.

- [61] 神原誠之, 岩尾友秀, 横矢直和. 拡張現実感のための動的シャドウマップを用いたシャドウレンダリング手法. 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 297–304, 2005.
- [62] S. Gibson, J. Cook, T. Howard, and R. Hubbard. Rapid shadow generation in real-world lighting environments. In *Proc. Eurographics Workshop on Rendering*, pp. 219–229, 2003.
- [63] A. Fournier, A.S. Gunawan, and C. Romanzin. Common illumination between real and computer generated scenes. In *Proc. Graphics Interface '93*, pp. 254–262, 1993.
- [64] M.F. Cohen, S.E. Chen, J.R. Wallace, and D.P. Greenberg. A progressive refinement approach to fast radiosity image generation. In *Proc. ACM SIGGRAPH '88*, pp. 75–84, 1988.
- [65] G.J. Ward. The RADIANCE lighting simulation and rendering system. In *Proc. ACM SIGGRAPH '94*, 1994.
- [66] 佐藤いまり, 林田守広, 甲斐郁代, 佐藤洋一, 池内克史. 実光源環境下での画像生成: 基礎画像の線形和による高速レンダリング手法. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 8, pp. 1864–1872, 2001.
- [67] J.S. Nimeroff, E. Simoncelli, and J. Dorsey. Efficient re-rendering of naturally illuminated environments. In *Proc. Eurographics Workshop on Rendering*, pp. 359–373, 1994.
- [68] 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.

- [69] 和田浩明, 池田聖, 佐藤智和, 横矢直和. 複数経路上を歩行移動可能なテレプレゼンスシステムにおける違和感の低減に関する研究. 日本バーチャルリアリティ学会サイバースペースと仮想都市研究会研究報告, 第11巻, pp. 29–34, 2006.
- [70] N. Kawai, K. Machikita, T. Sato, and N. Yokoya. Video completion for generating omnidirectional video without invisible areas. *IPSJ Trans. on Computer Vision and Applications*, Vol. 2, pp. 200–213, 2010.
- [71] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *Visualization and Computer Graphics, IEEE Transactions on*, Vol. 4, No. 4, pp. 293–306, 2002.
- [72] T. Sato, H. Koshizawa, and N. Yokoya. Omnidirectional free-viewpoint rendering using a deformable 3-d mesh model. *Int. J. of Virtual Reality*, Vol. 9, No. 1, pp. 37–44, 2010.
- [73] T. Sato and N. Yokoya. Efficient hundreds-baseline stereo by counting interest points for moving omni-directional multi-camera system. *J. of Visual Communication and Image Representation*, Vol. 21, No. 5–6, pp. 416–426, 2010.
- [74] M. Hori, M. Kanbara, and N. Yokoya. Arbitrary stereoscopic view generation using multiple omnidirectional image sequences. In *Proc. 20th IAPR Int. Conf. on Pattern Recognition (ICPR2010)*, pp. 286–289, 2010.
- [75] S. Ikeda, T. Sato, and N. Yokoya. High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system. In *Proc. 2003 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System (MFI2003)*, pp. 155–160, 2003.
- [76] N. Igawa, Y. Koga, T. Matsuzawa, and H. Nakamura. Models of sky radiance distribution and sky luminance distribution. *Solar Energy*, Vol. 77, pp. 137–157, 2004.

付録

A. 四元数および球面線形補間

本節では、全方位動画の姿勢統一で用いた四元数（クォータニオン：Quaternion）および四元数を用いた球面線形補間について述べる。

A.1 四元数を用いた回転表現

四元数を用いた表現は3次元空間内での回転を表現する手法の一つであり、4つの実数パラメータで任意の回転を表現する。Microsoft社のDirect3Dは、回転表現として四元数を用いている。オイラー角表現で問題となる、特定の軸周りの回転が一意に定義できなくなるジンバルロックが発生せず、回転行列と比較しパラメータが少ないことが利点である。本研究においては、次項で述べる球面線形補間を容易に実現できるため四元数を用いる。実部 t および虚部 x, y, z を用い、四元数 \mathbf{Q} を以下のように表す。

$$\mathbf{Q} = (t; x, y, z) \quad (8)$$

各パラメータは、回転軸と回転量から計算できる。回転軸を (α, β, γ) 、回転量を θ としたとき、四元数 \mathbf{Q} および共役四元数 \mathbf{Q}^{-1} は以下のように定義される。

$$\mathbf{Q} = \left(\cos\left(\frac{\theta}{2}\right); \alpha \sin\left(\frac{\theta}{2}\right), \beta \sin\left(\frac{\theta}{2}\right), \gamma \sin\left(\frac{\theta}{2}\right) \right) \quad (9)$$

$$\mathbf{Q}^{-1} = \left(\cos\left(\frac{\theta}{2}\right); -\alpha \sin\left(\frac{\theta}{2}\right), -\beta \sin\left(\frac{\theta}{2}\right), -\gamma \sin\left(\frac{\theta}{2}\right) \right) \quad (10)$$

また、四元数で表現された回転の合成には、以下に示す四元数の積を用いる。

$$\begin{aligned} \mathbf{Q} &= (s; \mathbf{q}) \\ \mathbf{R} &= (t; \mathbf{r}) \\ \mathbf{QR} &= (st - \mathbf{q} \cdot \mathbf{r}; s\mathbf{q} + t\mathbf{r} + \mathbf{q} \times \mathbf{r}) \end{aligned} \quad (11)$$

3次元空間中の点 $\mathbf{v} = (a, b, c)$ は四元数を用いて以下のように表現される.

$$\mathbf{V} = (0; \mathbf{v}) = (0; a, b, c) \quad (12)$$

3次元空間中のある点 \mathbf{v} に対して原点を中心とした回転を行った点 \mathbf{w} は以下のよう
に求められる.

$$\mathbf{Q}^{-1}\mathbf{V}\mathbf{Q} = (0; \mathbf{w}) \quad (13)$$

A.2 球面線形補間

球面線形補間は、回転の補間に多く用いられる補間手法である. 回転軸 \mathbf{c}_q , 回転量 ϕ_q で定義される回転 \mathbf{Q} および回転軸 \mathbf{c}_r , 回転量 ϕ_r で定義される回転 \mathbf{R} の間を滑らかに補間する場合, 回転軸および回転量の両方を滑らかに補間する必要がある. 球面線形補間は, 原点を中心とした回転を用いて球面上を沿うように補間を行うことにより, これを実現する. オイラー角表現等では, この補完を実現するための計算が複雑になる. しかし, 四元数を用いると, \mathbf{Q} から \mathbf{R} への回転を表す四元数 \mathbf{U} を求めた上で, $t = [0, 1]$ の媒介変数を用い $t\mathbf{U}$ を求めることで容易に補間可能である.