信学技報 TECHNICAL REPORT OF IEICE.

# フライスルー MR 平城京

無人飛行船からの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

## 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 情報科学研究科 〒 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**本報告では、無人飛行船に搭載された全方位カメラから撮影された動画像を用いた、広域屋外環境における蓄積再生型拡張テレプレゼンスシステムを提案する.提案システムは、あらかじめ撮影された空撮全方位動画像からカメラ位置・姿勢の推定を行い、さらに動画像中における飛行船の映り込み等の死角領域を補完し実環境の環境マップを生成し光源環境として用いることで、幾何学的・光学的整合性の実現を図る.本稿では、空撮動画像からのカメラ位置姿勢の推定手法および光源環境の推定手法を提案し、平城宮跡上空で撮影された動画像を用いた全方位 AR シーンの生成実験結果について述べる.

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

# Fly-through MR Heijo-kyo

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

Fumio OKURA<sup>†</sup>, Masayuki KANBARA<sup>†</sup>, and Naokazu YOKOYA<sup>†</sup>

 † Graduate School of Information Science, Nara Institute of Science and Technology (NAIST) 8916–5, Takayama, Ikoma, Nara, 630–0192 Japan E-mail: †{fumio-o,kanbara,yokoya}@is.naist.jp

**Abstract** This report proposes an augmented telepresence system which generates large-scale omnidirectional AR scenes in outdoor environments using recorded omnidirectional videos captured from an unmanned airship. The proposed system realizes geometric and photometric registration by estimating extrinsic camera parameters and illumination environment of real scene from an input omnidirectional video. This report describes a system for aerial imaging using an airship and an omnidirectional camera, geometric and photometric registration methods, and experimental results of generating AR scenes from an omnidirectional video sequence.

Key words augmented telepresence, aerial image, unmanned airship, geometric registration, photometric registration

## 1. はじめに

近年,ユーザに臨場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[1],[2].特に,遠 隔地のロボット等を用いた研究は,人による偵察の難しい災害 地等において有効であると考えられている[3].また,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,カメラ等で取得し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 [4], [5].本研究ではこれを拡張テ レプレゼンス (Augmented Telepresence) と呼ぶ.

本研究は地上での撮影と比較して広範囲の情景を取得するた め,飛行船と全方位カメラを用いて空撮された動画像に建造物 の三次元モデルを重畳する蓄積再生型の拡張テレプレゼンスシ ステムを構築する.本研究では蓄積再生型のシステムを扱うた め,幾何学的・光学的整合性をオフライン処理で実現する.幾 何学的整合性を実現するため全方位カメラの位置・姿勢を推定



図1 提案システムの流れ



図2 移動体別の空撮高度

する.また,光学的整合性を実現するため飛行船等の全方位動 画像への映り込み領域を補完することで実環境の環境マップを 生成し,光源環境として用いることで実物体と仮想物体間の光 源環境を統一する.

本稿では、2. 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・姿勢の推定法、光源環境の推定法について述べる. また、3. で平城宮跡上空から撮影された動画像を用いた全方位 AR シーンの生成実験およびその結果を示す.

## 2. 飛行船と全方位カメラによる拡張テレプレゼ ンス

#### 2.1 システムの処理概要

提案システムの流れを図1に示す.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(処理1),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(処理2).次に空撮動画像およびカメラ位置・姿勢を用 いて飛行船の全方位動画像への映り込み等を補完することで生 成された実環境の環境マップを光源環境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し,同時にユーザへ直接提示 することで飛行船等による風景への遮蔽に対処する(処理3). 最後に得られた撮影位置・姿勢情報付き全方位動画像および光 源環境を用いて市販のレンダリングソフトウェアによりグロー バルイルミネーション(GI)を用いたレンダリングを行うこと により,全方位 AR シーンの生成を行う(処理4).

以下,全方位空撮動画像の取得,カメラ位置・姿勢の推定手法,および実環境の環境マップ生成と光源環境への適用手法の 詳細を述べる.



図3 無人飛行船の外観と搭載機器

### 2.2 無人飛行船による全方位空撮動画像の取得

拡張テレプレゼンスにおいて,空撮は地上からの撮影と比較 して鳥瞰視点で広範囲の動画像を取得できるため,広域屋外環 境における拡張テレプレゼンスシステムを構築するために有効 な撮影手段である.特に,図2に示すように無人航空機(飛行 機,ヘリコプタ,飛行船等)は有人航空機と比較して低空での 撮影に適しており,建造物等を詳細に確認するような用途に有 効である.無人飛行船は飛行機やヘリコプタと比較すると安全 であり,さらに撮影を低コストで行うことができる.このよう な理由から,本研究では空撮に無人飛行船を用いる.

提案システムでは、図3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、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.飛行船の仕様を表1,各センサの詳細を表2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.

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4に示すようなパノラマ画像に変換される[6].

#### 2.3 全方位カメラの位置・姿勢推定

提案システムでは、蓄積誤差が発生する、スケールが未知で

表1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 (バッテリ駆動)連続航行時間約 60 分搭載センサGPS・姿勢センサ・全方位カメラ

表 2	搭載センサの仕様
-----	----------

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



図 4 全方位カメラで得られる空撮パノラマ画像

あるといった問題点が存在する structure-from-motion (SfM) アルゴリズムを改良し,蓄積誤差の発生しない GPS の位置情 報を用いて GPS の測地座標系におけるカメラ位置・姿勢を得 る手法 [7] を用いる.

ー般に 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 $E_{qij}$ を以下のように定義する.

$$E_{qij} = |\mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij}|^2 \tag{1}$$

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 $g_i$ 、カメラパラメータ行列を  $M_i$  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 d は一定となり、以下の関係が成立する.

$$\mathbf{M}_i \mathbf{g}_i = \mathbf{d} \tag{2}$$

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  $E_{gi}$ を以下のように定義する.

$$E_{qi} = |\mathbf{d} - \mathbf{M}_i \mathbf{g}_i|^2 \tag{3}$$

誤差関数 *E<sub>qij</sub>* および *E<sub>gi</sub>* を用い,評価関数 *E* は以下のように 定義される.

$$E = \frac{\omega}{|\mathbf{F}_{\mathbf{g}}|} \sum_{i \in \mathbf{F}_{\mathbf{g}}} E_{gi} + \frac{(1-\omega)}{\sum_{i} |\mathbf{S}_{\mathbf{p}_{i}}|} \sum_{i} \sum_{j \in \mathbf{S}_{\mathbf{p}_{i}}} E_{qij}$$
(4)

ここで、 $\omega$ は重み係数、 $\mathbf{S}_{\mathbf{p}_i}$ は第iフレームにおいて観測され た特徴点の集合、 $\mathbf{F}_{\mathbf{g}}$ 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・姿勢は、適切な $\omega$ を用いて評価関数 Eを勾配法により最小化することで求められる.

大量のフレームを一度に扱う場合,文献[7]の手法を用いる と,GPSの測位誤差に起因してトラッキングが破綻するため, 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.推定値が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.また,各動画像における1フレーム目の姿勢の初期値 には,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 を用いる.

#### 2.4 実環境の環境マップ推定

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,その領域の補完を行い実環境 の環境マップを生成することにより実環境の光源環境推定およ び地上等への遮蔽による実環境の視認性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.そのため,死角領域 は事前にマスクを作成することによって指定する.

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5のように飛行船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,環境マップ推定処理の流れを述べる.

- (1) 地上領域に存在する死角領域を補完する.
- (2) 動画像のみから空に存在する死角領域の輝度値を推定する.
- (3) 残った死角領域を天空光モデルを用いて補完する.
- (4) レンダリング時の照明環境として用いるため、実環境・ 仮想環境間における輝度比を用いて環境マップの補正 を行う.

以下,各処理の詳細を述べる.

2.4.1 地上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 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、領域の移動 ベクトルを求める.推定された移動ベクトルを用いて前フレー ムの対応領域をコピーし、地上付近の死角領域を補完する.こ れを初期フレームから繰り返すことにより、図6のように動画 像全体における地上付近の死角領域を補完する.

2.4.2 空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,動画像全体で空の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,動画像全体で1枚の空画像を生成する.また空は無限遠に存在するとして,2.4.1のような画像上での位置合わせを行わずに補完を行う.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.まず,2.3で推定されたカメラの姿勢を用いて方角を合わせ,図7のような空画像を生成する.図7には,死角領域としてマスクされ



図 5 飛行船の映り込みによる地上への遮蔽





- 図7 各フレームでの空画像 (北が上方の等距離射影画像)
- 図8 動画像全体を用いて 生成される統合後の空画像



図 6 地上に存在する死角領域の補完結果





図 9 CIE Overcast Sky Model 図 10 補完後の HDR 空画像 に基づいて生成される空画像

(表示輝度値は図7に統一)

た部分を塗りつぶし、北を上方に合わせた等距離射影画像を示 す.次に、各フレームを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 vxy は, 第 iフレームの空画像の輝度値 vixy を用いて以下のように計算さ れる.

$$v_{xy} = \frac{1}{\sum_{i} \alpha} \sum_{i} \alpha_i * v_{ixy} \tag{5}$$

統合後の空画像を図8に示す.

2.4.3 残った死角領域の天空光モデルを用いた補完

図8において塗りつぶされた部分のように、動画像を通して 全てのフレームで空が映り込まず死角となる部分では $\sum_{i} \alpha = 0$ となり、統合後の空画像の輝度値を決定できないため補完が必 要である.補完には、空全体の輝度を統計的にモデル化した天 空光モデルの一種であり曇天時の輝度分布を表す CIE Overcast Sky Model [8] (図 9 参照) を用いて,死角領域における各画素 の輝度値を計算し補完を行うことで、図10のような空画像を 生成する.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. Overcast Sky Model における天空の輝度は以下のように定式化される. ここ で、θは天頂からの角度であり、Yzは天頂の放射輝度を表す.

$$Y = Y_z \frac{1+2\cos\theta}{3} \tag{7}$$

提案システムにおける放射輝度は、測定を行っていないため未 知である.そこで、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計算を行 う.ここで、多くの場合天頂が死角となり、天頂の輝度値 Yz は



図 11 画像全体の補完結果例 (パノラマ)

未知となる. そこで, 式(7)より

$$Y_z = \frac{3Y}{1+2\cos\theta} \tag{8}$$

として, 死角領域でない, かつ輝度値の飽和が起こっていない 各画素の輝度値を Y としてそれぞれ Yz を求め、これらを平均 することで天頂の輝度値を得る. 天頂の輝度値を用い, 死角領 域における輝度値を式(7)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11のように動画像 全体を補完する. 仮想物体を重畳する際の背景画像として, 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.

2.4.4 光源環境として用いるための環境マップ補正

環境マップの空領域を用いて, Image Based Lighting (IBL) により光源環境を AR シーンに反映する.しかし,光源として 与えるべき輝度と環境マップの輝度値の対応は未知であるため、 このまま光源環境として用いるのは不適切である.そこで、推 定された環境マップをそのまま照明として用いてレンダリング された仮想物体上の画素と反射特性が等しいと考えられる実物 体上の画素の対応を手動で指定することにより両者の対応を求 め、補正する.対応画素はユーザによって画像上をクリックす るなどにより指定される.指定された面が完全拡散面であると 仮定すると、対応する画素を含む面の法線方向が同じであり、 かつ物体の反射特性が等しく、空気中の減衰が無視できる場合、 対応画素の輝度値の差異は光源環境の違いのみに起因する.こ のとき、仮想物体上の画素の輝度値  $I_v$  および実物体上の画素 の輝度値  $I_r$  の関係は以下のように近似できる.

$$I_r = \alpha I_v \tag{9}$$

比例係数  $\alpha$  を画像全体で 1 つ求め,環境マップによる照明強度 として用いることで補正を行う.  $\alpha$  は最低 1 組の  $I_v$ ,  $I_r$  の対 応で求められるが,複数の対応が与えられた場合には,各対応 によって求められた  $\alpha$  の平均値をパラメータとして用いる. 仮 想物体の反射係数が,必ずしも実物体と一致した設計となって いないため,通常は複数の対応が与えられるべきである.

### **2.5** 全方位 AR シーンの生成・提示

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..

生成・蓄積された全方位 AR シーンは,既存の全方位動画像 を用いたテレプレゼンスシステム [9] 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変換され,HMD や大型ディス プレイ等のディスプレイデバイスを用いて提示される.

## 3. 実 験

提案システムの有効性を確認するため、無人飛行船を用いて 平城宮跡上空で撮影を行ない、仮想物体を重畳する実験を行っ 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.飛行船に搭載された GPS によって得 られた飛行パスを図 12 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 程度であった.重畳した仮想物体 には、図 13 に示される平城宮の三次元モデルのうち、建物部 分のみを用いた.

生成された全方位 AR シーンの例を図 14 に示す.また,全 方位動画像を用いたテレプレゼンスシステム [9] を用いて全方 位画像から変換された平面透視投影画像を図 15 に示す. 概ね 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.



135.7945 135.795 135.7955 135.796 135.7965 135.797 135.7975 経度

図 12 平城宮跡上空からの撮影時における飛行船の軌跡 △: シーケンス開始地点, ○: シーケンス終了地点)



図 13 平城宮の三次元モデル(凸版印刷株式会社製作)

### 4. ま と め

本稿では,幾何学的・光学的整合性を考慮してレンダリング を行うことで,臨場感の高いARシーンを生成・提示する広域 屋外環境における蓄積再生型拡張テレプレゼンスシステムを提 案した.提案システムでは動画像とGPSを用いてカメラ位置・ 姿勢を推定することで位置合わせを行い,全方位カメラ中の死 角領域を補完することで光源環境を推定してレンダリング時に 利用し,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像を用いたシステムを構築するこ とにより提案システムにおける仮想物体の位置合わせおよび光 源環境の統一,全方位ARシーンの生成の有効性を確認した.

今後は、光源環境の推定等を実時間処理に対応させ、実時間 拡張テレプレゼンスに応用していく予定である.

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.19200016)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

S. Moezzi. Special issue on immersive telepresence. *IEEE Multimedia*, Vol. 4, No. 1, pp. 17–56, 1997.



(a) 入力された全方位画像

(b) 生成された全方位 AR 画像

図 14 提案システムにより生成された全方位 AR シーン:カメラの姿勢情報を用いて方角を調整後、メルカトル図法で展開



図 15 全方位 AR シーンから変換された平面透視投影画像の例

- [2] M. Hori, M. Kanbara, and N. Yokoya. MR telepresence system with inertial force sensation using a motion platform and an immersive display. In *Proc. IEEE Symposium on* 3D User Interfaces (3DUI), pp. 133–134, 2009.
- [3] T. Fong and C. Thorpe. Vehicle teleoperation interfaces. Autonomous Robots, Vol. 11, pp. 9–18, 2001.
- [4] 大倉史生,神原誠之,横矢直和.空撮画像を用いた拡張テレプレゼンス-無人飛行船の自動操縦と全方位カメラによる AR システム-. 画像の認識・理解シンポジウム (MIRU2010) 講演論文集, pp. 1183–1189, 2010.
- [5] F. Okura, M. Kanbara, and N. Yokoya. Augmented telepresence using autopilot airship and omnmi-directional camera. In Proc. 9th IEEE Int. Sympo. on Mixed and Augmented Reality (ISMAR 10), 2010. (To appear).
- [6] S. Ikeda, T. Sato, and N. Yokoya. High-resolution

panoramic movie generation from video streams acquired by an omnidirectional multi-camera system. In *Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System (MFI2003)*, pp. 155–160, 2003.

- [7] Y. Yokochi, S. Ikeda, T. Sato, and N. Yokoya. Extrinsic camera parameter estimation based-on feature tracking and GPS data. In Proc. 7th Asian Conf. on Computer Vision (ACCV2006), Vol. 1, pp. 369–378, 2006.
- [8] CIE-110-1994. Spacial distribution of daylight luminance distributions of various reference skies. In International Commission on Illumination, 1994.
- [9] 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.