

全方位型マルチカメラシステムを用いた高臨場感 テレプレゼンスシステムの構築手法に関する研究

池田 聖 横矢 直和

奈良先端科学技術大学院大学 情報科学研究科

あらまし: 視点位置や視線方向に応じた実世界映像の実時間提示によるテレプレゼンスでは, 実環境におけるユーザの視界を忠実に再現し, 高い臨場感を与えることが課題とされる. 本研究では, 従来の一般的なカメラよりも高解像度, 広視野角な画像を取得可能な全方位型マルチカメラシステムの導入により, テレプレゼンスシステム利用時のユーザの臨場感の向上を図る. 本報告では, 全方位型マルチカメラシステムを用いて提示画像を生成するために必要なキャリブレーション手法を提案し, テレプレゼンスシステムのプロトタイプを試作することにより提案手法の有効性を確認する.

Construction of Telepresence Systems Using an Omnidirectional Multi-camera System

Sei IKEDA and Naokazu YOKOYA

Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: One of the most important problems in telepresence is to improve the quality of images presented to users in order to provide a rich sense of presence. The approach for this problem in this study is based on employing an omnidirectional multi-camera system which can obtain high resolution and wide field-of-view videos. This report describes novel calibration methods for an omnidirectional multi-camera system and shows two prototype telepresence systems to confirm the validity of the proposed methods.

1 序論

実画像提示により遠隔地の臨場感を再現するテレプレゼンスは, エンターテインメントや教育, 医療などへの応用が期待されている. テレプレゼンスでは, 要求される視点位置や視線方向に応じた実画像の実時間提示により実環境における視界を再現することでユーザに臨場感を与える. テレプレゼンスシステムは, ユーザの仮想視点を能動的に移動する方式と受動的に移動する方式に大別されるが, 両方式共に, ユーザに高い臨場感を与えることが課題とされている. 本研究では, 受動的・能動的視点移動型の両方式のテレプレゼンスシステムにおいてユーザに高い臨場感を与えるために, 高解像度, 広視野角な全方位型マルチカメラシステムを用いて生成した全天球動画を用いる. 特に, 能動的視点移動型のテレプレゼンスシステムに対しては, 上記の両方式に共通する手法に加えて, 歩行動作インタフェースを用いた画像提示システムを用い, 利用時の歩行動作による頭部位置の変動および歩行動作と相関の無い撮影時の全方位型マルチカメラシステムの移動

速度や姿勢の変動を考慮した画像提示により臨場感の向上を図る.

全方位型マルチカメラシステムにより得られる複数の異なる方向の画像からテレプレゼンスにおいて提示される画像を生成するためには, カメラシステムの内部パラメータが高精度に推定されている必要がある. 従来の研究 [1, 2, 3] では, 内部パラメータの一部にカメラシステムの設計値を用いることを前提としており, これが入手できない場合に問題となる. 本研究では, 設計値を用いることなく, 使用するマーカの配置法の改良等により高精度な内部パラメータの推定を試みる.

また, 撮影時のカメラシステムの移動速度や姿勢の変動を考慮した画像提示を行うためには, カメラシステムの外部パラメータの推定が必要となる. 広範囲な視点移動を再現するテレプレゼンスに利用するためには, 蓄積誤差の発生を回避しなければならない. 従来研究では, この問題に対して, GPS, 磁力計, 2軸角度センサーを組み合わせる手法 [4] や画像上の特徴追跡に加えて環境の三次元的な幾何情報を利用する手法 [5, 6, 7, 8] が提案されている. しかし, 前者の手法では, センサ間

のキャリブレーションの問題や GPS の位置情報とカメラシステムの動画間の取得レート違いを解決する必要があり、後者の手法では、幾何情報の取得に人手を要するという問題がある。本研究では、画像上の特徴追跡に加えて GPS の位置情報を利用することにより、蓄積誤差の発生を回避し、精度の高い外部パラメータの推定を試みる。

以下、2章では、全方位型マルチカメラシステムの内部パラメータの推定法およびについて述べ、推定されたカメラ内部パラメータを基に生成された全天球画像を用いる受動的視点移動型テレプレゼンスシステムの試作結果を示す。3章では、全方位型マルチカメラシステムの外部パラメータの推定法について述べる。4章では、これらの全方位型マルチカメラシステムの内部・外部パラメータに基づく画像提示法および歩行動作インタフェースを用いた能動的視点移動型テレプレゼンスシステムについて述べる。最後に5章で、全体を総括し、今後の展望について述べる。

2 全方位型マルチカメラシステムのカメラ内部パラメータ推定

2.1 緒言

全方位型マルチカメラシステムの内部パラメータとは、カメラシステムの内部構造を表す変数であり、撮影中に変化するカメラの位置・姿勢に対して、撮影時に変化しないものを意味する。複数のカメラの位置・姿勢関係および個々のカメラの入力画像の歪みや焦点距離を表すパラメータを推定する幾何学的キャリブレーションと、個々のカメラの明度低下現象を補正し、カメラ間の相対的な色調に関するパラメータを推定する光学的キャリブレーションがある。本研究では、これらのパラメータは撮影中は変化しないものと仮定し、全てのパラメータを予め専用パターン等を用いて推定しておくことで、利用時に取得する画像の内容に依存せず一定の提示画像の生成や外部パラメータの推定に利用することを考える。

ただし、本報告では、全方位型マルチカメラシステムの内部・外部パラメータと、一般的なカメラの内部・外部パラメータを区別するために、一般的なカメラのパラメータは、「単眼カメラの内部パラメータ」および「単眼カメラの外部パラメータ」と表記する。

2.2 全方位型マルチカメラシステム： Ladybug

全方位型マルチカメラシステムの例として本研究で使用する Point Grey Research 社製の Ladybug [9] について説明する。Ladybug は図 1(左) に示すような水平方向に 5 個、上方向に 1 個の CCD カメラを外向きに



図 1: Ladybug のカメラヘッド (左) と蓄積ユニット (右)



図 2: 取得画像 (上方向 (右下) と水平方向 (その他))

配置した小型カメラヘッドと図 1(右) に示すようなハードディスク 4 台からなる HDD アレイによる蓄積ユニットにより構成される。Ladybug は図 2 のような横 768 画素、縦 1,024 画素の高解像度な画像を 6 台のカメラから同期して取得し、上方を含む全方向の約 75%以上を 15fps の動画像として約 20 分間蓄積することが可能なシステムである。カメラ間の投影中心は一致しておらず、約 4cm の間隔がある。

2.3 幾何学的キャリブレーション法

幾何学的キャリブレーションでは、全方位型マルチカメラシステムを構成する各カメラに対して一般的な単眼カメラの内部パラメータおよび外部パラメータパラメータの推定を行う。一般的なカメラのキャリブレーションでは、単眼カメラの内部パラメータに相当する焦点距離 f 、歪み係数 $(\kappa_1, \kappa_2, \kappa_3)$ 、歪み中心 (c_x, c_y) 、アスペクト比 s_x と単眼カメラの外部パラメータに相当するカメラ c の位置 T_c 、姿勢 R_c を推定しなければならない。

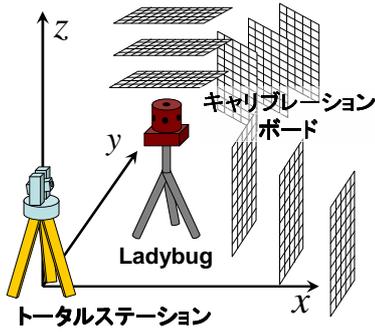


図 3: キャリブレーションボードの配置

マーカの配置

上述のカメラパラメータを正確に推定するためには、カメラシステムの全方向に三次元位置関係が既知のマーカを多数配置する必要がある。本手法では、マーカとしてキャリブレーションボード上の格子模様の交点を用いる。マーカの三次元位置は、キャリブレーションボードの3隅の格子点をトータルステーションで計測し線形補間して求める。キャリブレーションボードは図3に示すように、各カメラに対して奥行き方向に移動させ、その都度三次元位置の計測と、キャリブレーションボードの撮影を行なう。これにより、容易に多数のマーカを空間的に配置することが可能となる。全てのカメラでキャリブレーションボードを撮影し終わるまで、トータルステーションにより張られる世界座標系を統一することで、カメラシステムの全方向にマーカを同時に配置するのと同等のデータを得ることが可能である。

幾何学的パラメータの推定

各カメラの幾何学的パラメータの推定方法について述べる。単眼カメラの内部パラメータはあらかじめ各カメラについて Tsai [10] の手法を用いて推定しておく。ただし、歪み係数については3次係数 κ_3 まで考慮し精度向上を図る。

単眼カメラの外部パラメータは線形演算と非線形最適化により画像上の再投影誤差の二乗和を最小化することで求める [11]。本手法では、単眼カメラの外部パラメータは、世界座標系におけるカメラ $c(c = 0, 1, \dots, 5)$ の位置 \mathbf{T}_c と姿勢 R_c を用いて、世界座標系からカメラ座標系への変換行列 M_c として表わされる。

$$M_c = \begin{bmatrix} R_c & \mathbf{T}_c \\ 0 & 1 \end{bmatrix} \quad (1)$$

M_c を求めるために、マーカ $m(m = 1, 2, \dots)$ の三次元位置 \mathbf{x}_m とカメラ $c(c = 0, 1, \dots, 5)$ の入力画像上での位置 \mathbf{u}_m とを対応づけ、線形解法 [12] を用いて世界座標系からカメラ座標系への変換行列 M'_c を計算する。しかし

M'_c の自由度は12となり回転行列 R_c が正規直交基底を成さない。そこで、真のカメラ位置が推定された光軸上に存在すると仮定して、線形演算により6自由度に補正された変換行列 M_c を得る。最後に、マーカの三次元位置を画像上に投影した座標 \mathbf{v}_m とマーカの入力画像上での位置 \mathbf{u}_m としたときに、以下のように表される再投影誤差の二乗和を評価関数として勾配法によって最小化することで変換行列 M_c を更新する。

$$E_c = \sum_m |\mathbf{u}_m - \mathbf{v}_m|^2 \quad (2)$$

2.4 光学的キャリブレーション法

明度低下現象の補正

一般的な全方位型マルチカメラシステムでは、広角レンズを用いて広い視野の画像を取得するため、画像の周辺部分が暗くなる明度低下現象が生じる。明度低下現象としては、広角レンズ特有の $\cos^4 \theta$ 現象 [13] と望遠レンズに特有の口径蝕現象 [14] が知られているが、広角レンズを用いたレンズ系では前者の影響が支配的であるため、ここでは $\cos^4 \theta$ 現象のみを考慮する。

Horn [13] によれば、直径 l のレンズへ光軸に対して入射角 θ の角度をとる方向から入射する光の殆どがレンズの中心を通過することを仮定すれば、レンズへ入射する光の強度 I' に対する撮像面上で放射照度 I は、レンズへ入射する光の入射角 θ に依存して減衰し、以下の式を満たす。

$$I = \frac{\pi l^2 \cos^4 \theta}{4f^2} I' \quad (3)$$

本研究ではレンズ径の実測値 l と幾何学的キャリブレーションにおいて推定される焦点距離 f を使用し、上の式を用いることで $\cos^4 \theta$ 現象による明度低下を補正する。

カメラ間の相対的な色調補正

一般に撮像面上の放射照度 I は物体表面の放射輝度 L に対して線形的な関係 $I = aL' + b$ を持つとされており、異なるカメラ c と c' における撮像面上の放射照度 I_c と $I_{c'}$ の間にも以下に示す線形的な関係があると考えられる。

$$I_{c'} = a_c I_c + b_c \quad (4)$$

これらの線形変換係数 a_c, b_c が各カメラごとに異なるためカメラ間での色調の違いが発生する。本研究では、色調のモデルとして上式で示されるモデルを用い、各カメラ c の変換係数 a_c, b_c を推定するために以下のような手順をとる。

1. 照明条件を固定した同じシーンを撮影し、RGB それぞれの輝度値についてヒストグラムを作成する。

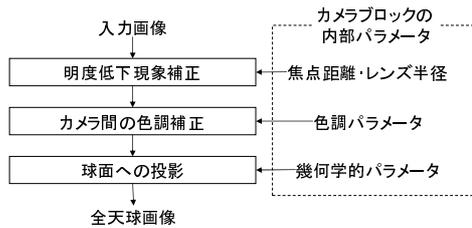


図 4: 全天球画像の生成手順

2. カメラ c の画像と基準カメラ c_0 の画像の輝度 i に関するヒストグラム $h_c(i)$ と $h_{c_0}(i)$ が一致するように、以下の評価関数を最小化し、各カメラの a_c, b_c を推定する.

$$e(a_c, b_c) = \sum_i \left\{ h_{c_0}(i) - \frac{1}{a_c} h_c \left(\frac{i - b_c}{a_c} \right) \right\}^2 \quad (5)$$

3. 推定したパラメータから式 (4) に基づき、入力画像の RGB の輝度値をそれぞれ独立に変換する.

2.5 全天球画像の生成手法

推定された全方位型マルチカメラシステムの内部パラメータを基に、入力画像から球面透視投影画像として全天球画像を生成する手法について述べる。図 4 に示すように、まず、前節で述べた入力画像の明度低下現象の補正およびカメラ間の色調補正を行い、次に仮想的な球面を設定し補正された入力画像を投影する。最後に、球面上に投影された画像を正距円筒図法により平面に展開する。以下では、仮想的な球面への画像の投影について述べる。

一般的に、カメラを放射状に配置するマルチカメラシステムでは、各カメラの投影中心は一致せず、視差が発生するために厳密な一点透視投影に基づく継ぎ目のない全天球画像を作ることは不可能である。しかし、撮影対象までの距離がカメラの投影中心間の距離に比べて十分大きい場合は、視差の影響が相対的に小さくなるため、各カメラの投影中心が一致していると見なすことができる。ここでは、撮影対象は十分遠方に存在するものと仮定し、投影面である球面 S の半径を十分に大きく設定する。球面の中心は全てのカメラの投影中心の重心位置 G に設定する。このような仮定の下で、各画像上の全ての画素を球面 S に投影することで全天球画像を生成する。ただし、球面 S 上の点 s の画素 $I_S(s)$ に対応する入力画像が複数存在する場合、複数の画像が滑らかにつながるように画像間のブレンド処理を行う。

以上の手法により、一点透視投影画像と同等の画像が得られる条件について述べる。図 5 に示すように撮影対象 x が 2 つのカメラ c, c' 上の点 $u_c, u_{c'}$ に撮影されたとき上述の手法では、球面 S 上の $s_c, s_{c'}$ に投影され、2 重

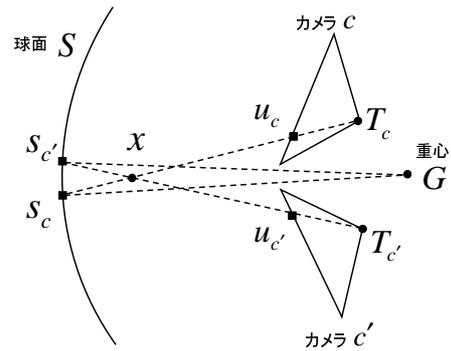


図 5: 視差の影響

にずれた画像が生成される。球面 S の円周長が N 画素で構成されているとすると、 $\angle s_c G s_{c'} < \frac{2\pi}{N}$ であれば、視差の影響は 1 画素以内に収まり、一点透視投影画像と同等の画像が得られる。無限遠に設定された球面を考えたとき $\angle s_c G s_{c'} \rightarrow \angle s_c x s_{c'}$ となるので、 $\angle T_c x T_{c'} < \frac{2\pi}{N}$ という条件を満たす範囲に撮影対象 x が存在すればよい。

2 台のカメラの投影中心間の距離を d とし、カメラから撮影対象 x までの距離が等しい場合に限定すると、視差の影響が 1 画素以下となるカメラ間のベースラインから撮影対象までの距離 λ は以下の式で表現できる。

$$\lambda > \frac{d}{2 \tan \frac{\pi}{N}} \quad (6)$$

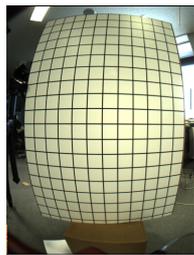
例えば、Ladybug の場合、隣り合うカメラの投影中心間の距離を 40mm、球面へ投影された画像の円周長が 3,340 画素で構成されていると仮定すると、視差の影響が 1 画素以内に収まる距離 λ は、約 20m である。

2.6 実験

2.6.1 幾何学的キャリブレーション

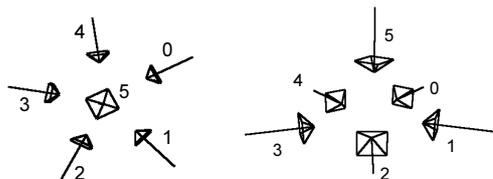
Ladybug の幾何学的なキャリブレーションでは、Ladybug を三脚に固定し、キャリブレーションボードをカメラの奥行き方向に約 50cm 間隔で移動させ、各カメラごとに 3 回づつ撮影した。水平方向の各カメラにはキャリブレーションボードの格子点のうちそれぞれ 187 個ずつ検出し、合計 561 個のマーカの三次元位置と画像上の位置を取得した。上方向のカメラについては、キャリブレーションボードを固定する部分がマーカを遮るため、各画像から 170 個ずつ検出し、合計 510 組のマーカの位置情報を取得した。マーカの三次元位置はキャリブレーションボードの 3 隅の格子点を LEICA 社製のトータルステーション TCR1105 XR で計測し、その他の格子点の三次元位置は線形補間により算出した。

まず、図 6 に推定した内部パラメータを用いた入力画像の歪み補正結果を示す。同図より格子模様が直線に



(a) レンズ歪み補正前 (b) レンズ歪み補正後

図 6: レンズ歪み補正結果



(a) 上からみた配置 (b) 正面からみた配置

図 7: カメラの位置および姿勢の推定結果

戻っていることから、おおむね正しく内部パラメータが推定されていることが確認できる。

次に、推定した外部パラメータを用いて描画したカメラの位置と姿勢の関係を図 7 に示す。同図では、四角錐と直線でカメラの位置と姿勢の関係を表現している。四角錐の頂点と直線が交わる点はカメラの投影中心を、直線の方向はカメラの光軸を、四角錐の底面はカメラの光軸周りの回転を表す。また、外部パラメータ最適化後の平均再投影誤差は、最も大きいカメラで 0.65 画素であった。図 7 に示した図形の位置および姿勢の関係が Ladybug のカメラの配置に近く、カメラの再投影誤差の平均値が小さいことから、推定した外部パラメータに大きな誤差が含まれていないことが確認できる。

2.6.2 光学的キャリブレーション

図 8 に明度低下現象の補正結果を示す。同図左のように補正前の画像では、周辺部分の輝度値が全体的に低下しているのに対して、同図右の様に補正後の画像ではそれらが補正されていることが分かる。

図 9 は、色調補正前後でブレンド処理なしに全天球画像を生成し、異なるカメラによる画像の継ぎ目部分を比較したものである。同図から色調補正前に比べて補正後の生成画像では、画像の境界が目立たなくなっていることが確認できる。

2.6.3 全天球動画の生成

推定した内部パラメータを用いて全天球動画を作成した。入力動画は、Ladybug をジェットコースターの上に固定し、走行中の景色を撮影することにより、2.2 節で図 2 に示した画像を得た。また、これらの入力画像



(a) 補正前の入力画像 (b) 補正後の入力画像

図 8: 明度低下現象の補正結果



(a) 色調補正していない画像 (b) 色調補した画像

図 9: 色調補正の効果

から作成した全天球画像を図 10 に示す。図 10 の下部の黒い部分は、入力画像の存在しない部分であり、全天球画像の最大の水平方向の解像度は、各カメラの水平方向の解像度である 768 画素を 5 倍したもから重複領域として水平解像度の 13% を差し引いた 3,340 画素とした。また、全天球画像の垂直方向の解像度は、水平方向の半分の 1,670pixel とした。図 10 から、全天球画像での位置ずれや入力画像間の境界は目立ず、幾何学的にも光学的にもおおむね正しくキャリブレーションが行なえていることが分かる。

また、異なるカメラの画像が重複して得られる領域にマーカを写し、全天球画像上でのマーカ位置間のずれを測定した結果、誤差 ϵ は平均 0.0061rad であった。これは入力画像面上では約 3 画素に相当する。しかし、全天球動画を表示し、約 3 画素の誤差が生じている部分を目視した結果、ブレンド処理の効果によりずれは認知できなかった。

2.6.4 受動的視点移動型テレプレゼンスシステムの試作

高解像度な全天球動画を用いた、図 11 に示す全天球の見回しが可能な受動的視点移動型テレプレゼンスシステムを試作した。試作システムは、全天球視野の半分程度を一度に表示可能な球面ディスプレイ (Elumens VisionStation)、コントローラ (Microsoft SideWinder Game Pad Pro)、PC (CPU: Intel Pentium4 1.7GHz、メモリ: 1GB、グラフィクスカード: Nvidia GeForce4) の 3 つの装置により構成されている。球面ディスプレ



図 10: 全天球動画画像の 1 フレーム



図 11: テレプレゼンスシステムの外観

この解像度が $1,024 \times 768$ であるため、全天球動画画像は、 $2,048 \times 1,024$ の解像度に縮小し、JPEG 形式で PC に蓄積してある。図 11 に示すように、利用者はコントローラを用いてインタラクティブに全方向の見回しや再生速度の変更、表示シーン切り替えを行うことができる。本システムでは Ladybug の撮影時のフレームレートである 15fps での描画が可能であった。

試作したテレプレゼンスシステムは解像度、視野角、対話性の面において臨場感の高いシステムとなったが、球面ディスプレイの解像度の制限のため入力画像の解像度を十分に生かしたものではない。より解像度の高いディスプレイを利用することで、さらに臨場感の高いシステムの構築が可能である。

2.7 結言

本章では、全天球型マルチカメラシステムの内部パラメータ推定として、幾何学および光学的なキャリブレーションを行い、その結果を基にして全天球動画画像を生成する手法を提案した。幾何学的キャリブレーションではキャリブレーションボードとトータルステーションを用いて、システムの全方向に多数のマーカを空間的に広く配置することで、設計パラメータを使用せずにカメラパラメータの推定を行った。本報告では Ladybug という特定のカメラシステムへの適用について述べたが、

提案したカメラ内部パラメータの推定手法は、放射状にカメラを配置した全天球型マルチカメラシステム全般に適用可能である。

また、全天球画像生成における画像のつなぎ合わせ精度を定量的に評価し、提案したカメラ内部パラメータ推定手法では、カメラから十分遠くにある点については入力画像面上で平均約 3 画素のずれが生じる精度でキャリブレーションを行えることが分かった。さらに、実際に全方向の見回しが可能なテレプレゼンスシステムを試作し、本研究で生成した全天球画像が受動的視点移動型テレプレゼンスに利用可能であることを確認した。

3 全方位型マルチカメラシステムのカメラ外部パラメータの推定

3.1 緒言

誤差が蓄積することなく、精度の高い外部パラメータを推定するためには、事前知識の獲得やセンサ間のキャリブレーションが必要という問題があった。本章では、屋外環境下を広範囲に移動する全方位型マルチカメラシステムで取得した動画画像からのカメラ外部パラメータの推定手法を提案する。提案手法では、動画画像中の特徴点の追跡と GPS 測位値を用いた最適化に基づきカメラ外部パラメータを推定する。特徴点の追跡では、ランドマーク・特徴点追跡によるカメラ外部パラメータの推定手法 [15] と同様に、特徴点の仮対応づけから推定される暫定的な外部パラメータを用いて特徴点の画像上での探索範囲を限定し、誤追跡を防ぐ。GPS 測位値を用いた最適化では、動画画像から推定される外部パラメータに基づく特徴点の再投影誤差および外部パラメータから推定される GPS 受信機の位置の推定誤差からなる評価関数を最小化することで誤差の蓄積を防ぐ。本手法では、現実環境に関する事前知識や GPS 受信機とカメラの間の正確なキャリブレーションを必要

とせず，GPS の測位座標系における外部パラメータを推定できる特長がある．

ただし，本手法では，動画像全体にわたりカメラに対する GPS 受信機の位置は固定されているものとし，容易に計測できる両センサ間の距離については既知，カメラ座標系における GPS 受信機の方位については推定するものとする．また，前章で述べた手法によりカメラ内部パラメータは推定されているものとし，本章で扱う画像は予め，歪み補正，周辺減光補正，カメラ間の色調補正がされているものとする．動画像と GPS 測位値の同期についても得られているものとする．

以下，3.2 節では，カメラと GPS 受信機の幾何学的関係のモデル化を行い，3.3 節では，提案手法の流れおよび各処理について詳述する．3.4 節では提案手法の有効性を確認するためにシミュレーションによる実験および実環境のデータを用いた実験を示し，最後に 3.5 節で本章をまとめる．

3.2 カメラと GPS 受信機の幾何学的モデル

本節では，まず，全方位型マルチカメラシステムのカメラ外部パラメータに関するカメラモデルについて述べ，次に従来の特徴点のみによるカメラパラメータ推定で一般的に用いられる再投影誤差について説明する．さらに，カメラと GPS 受信機の幾何学的関係をモデル化し，GPS 受信機に関する誤差について説明する．

全方位型マルチカメラシステムの外部パラメータ

本報告では，全方位型マルチカメラシステムの各カメラ間の位置および姿勢関係は，常に固定されていることを仮定しているため，全てのカメラの位置および姿勢は，代表するカメラ $c = 0$ の位置および姿勢により表現することができる．前章で述べたように，全方位型マルチカメラシステムの内部パラメータ推定により，各カメラ c のキャリブレーション時の世界座標系からカメラ座標系への変換行列 M_c が得られる．これを用いると，新たに撮影する環境の第 i フレームにおける世界座標系から各カメラ c のカメラ座標系への変換行列 N_{ic} は，以下のように表現できる．

$$N_{ic} = M_c(M_0)^{-1}N_{i0} \quad (c = 0, 1, 2, \dots) \quad (7)$$

$$= \begin{bmatrix} R_{ic} & \mathbf{t}_{ic} \\ 0 & 1 \end{bmatrix} \quad (8)$$

ただし， R_{ic} は，第 i フレームにおける世界座標系から各カメラ c の座標系への回転を表し， \mathbf{t}_{ic} は，並進を表す．このように考えることで，一般的な単眼のカメラと同様な式により全方位型マルチカメラシステムのカメラ

外部パラメータ ($R_i = R_{i0}$, $\mathbf{t}_i = \mathbf{t}_{i0}$) を表現することができる．

特徴点の再投影誤差

再投影誤差は，特徴点追跡に基づく外部パラメータ推定手法で一般的に用いられる誤差関数である [16, 17, 6]．再投影誤差の二乗和を最小化することでカメラパラメータを推定する手法はバンドル調整法と呼ばれている [18]．カメラ c に写る特徴点 j の第 i フレーム上での再投影誤差 Φ_{ijc} は，推定された三次元位置の画像上への投影座標 $\hat{\mathbf{q}}_{ijc}$ と画像上で検出される座標 \mathbf{q}_{ijc} を用いて，以下のように表される．

$$\Phi_{ijc} = |\mathbf{q}_{ijc} - \hat{\mathbf{q}}_{ijc}| \quad (j \in S_{ic}) \quad (9)$$

ただし， S_{ic} は第 i フレームにおいてカメラ c の画像上で検出される特徴点の集合を表す．

GPS 受信機とカメラに関する位置誤差

GPS の測位誤差には，真の GPS 受信機の位置を中心に発生する期待値 0 の局所的な測位誤差と平均観測位置が真値から外れる大域的な誤差の 2 種類が考えられる．本報告では，前者の局所的な測位誤差のみを考慮する．

一般に，GPS 測位値と外部パラメータの推定結果に誤差が含まれない場合，第 i フレームの外部パラメータ R_i , \mathbf{t}_i とそれに対応する GPS 測位値 \mathbf{g}_i および，カメラ座標系での GPS 受信機の位置 \mathbf{d} には，以下の関係が成り立つ．

$$R_i \mathbf{g}_i + \mathbf{t}_i = \mathbf{d} \quad (i \in \mathcal{F}) \quad (10)$$

ただし， \mathcal{F} は，GPS の計測時刻に対応する動画像のフレームの集合を表す．ここで，GPS 測位値 \mathbf{g}_i および外部パラメータ R_i , \mathbf{t}_i に誤差があると，式 (10) は成り立たない．そこで本研究では，外部パラメータ R_i , \mathbf{t}_i から推定される GPS 受信機の位置と GPS 測位値との差を GPS 受信機とカメラに関する位置誤差関数 Ψ_i とし，以下のように表す．

$$\Psi_i = |R_i \mathbf{g}_i + \mathbf{t}_i - \mathbf{d}| \quad (11)$$

3.3 GPS 測位値を用いた外部パラメータ推定

提案手法では，計算量の問題から従来手法による一括処理が難しいような広範囲にカメラが移動する環境を想定し，動画像と GPS 測位値から，特徴点の三次元位置，カメラ座標系における GPS 受信機の方位，カメラ外部パラメータを推定する．本手法は，主に特徴点追跡と外部パラメータの最適化処理からなり，両処理

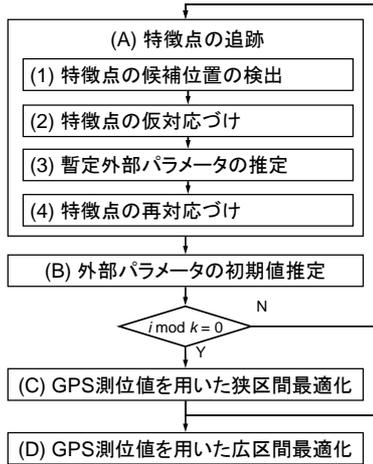


図 12: 外部パラメータ推定手法の処理手順

に GPS の位置情報を反映させることで、誤差の蓄積を防止する。

本手法の流れを図 12 に示す。まずフレーム毎の処理として、(A) 特徴点の追跡、(B) カメラ外部パラメータの初期値推定を行い、次に一定フレーム間隔 k で (C) GPS 測位値を用いた最適化を行う。処理 (C) では、蓄積誤差の影響によって後続のフレームで特徴点の対応付けの精度が低下することを防ぐために狭区間での最適化を行う。処理 (D) では、処理 (C) よりも広い区間で多くの GPS 測位値を用いて外部パラメータの最適化を行う。以下では、まず処理 (C)、(D) で共通に用いる GPS 受信機に関する誤差と再投影誤差の両方を考慮した誤差関数を定義し、その最適化手法について述べる。次に処理 (A) ~ (D) についてそれぞれ詳述する。

3.3.1 特徴点と GPS による誤差の定義とその最適化

ここでは、式 (9) で示した画像上の再投影誤差 Φ_{ijc} および式 (11) の GPS 受信機とカメラに関する位置誤差 Ψ_i を用いて外部パラメータ推定に用いる最適化の誤差関数 E を以下のように定義する。

$$E = \frac{\omega}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} \Psi_i^2 \quad (12)$$

$$+ \frac{1}{\sum_{i,c} |S_{ic}|} \sum_i \sum_c \mu_i \sum_{j \in S_{ic}} w_j \Phi_{ijc}^2 \quad (13)$$

ただし、 w_j は特徴点追跡処理 (A) で得た特徴点 j の信頼度であり、全てのフレームでの再投影誤差の分散の逆数とする。 μ_i は各フレームに対する重み係数を表す。 ω は Ψ_i および Φ_{ijc} に対する重み係数であり、予め、特徴点の検出位置精度および GPS の測位精度を考慮し、後述するシミュレーションにより、真値に対する誤差が最小になるような ω を算出しておく。一般に、シミュレ-

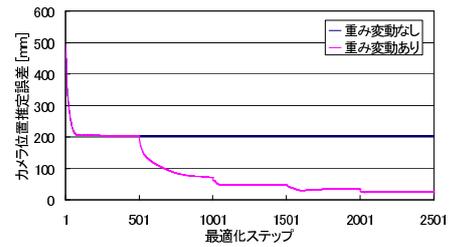


図 13: 重み変動の効果

ションと実利用時の環境では、GPS の測位回数 $|\mathcal{F}_i|$ と特徴点数 $\sum_i |S_{ic}|$ が異なるので、本研究では、これらの値に依らず重み係数 ω を一定値に決めるために、 $|\mathcal{F}_i|$ および $\sum_i |S_{ic}|$ により右辺の各項を正規化する。

式 (13) の誤差関数 E を用いた最適化では、外部パラメータ R_i , t_i 、特徴点の三次元位置 p_j 、およびカメラ座標系における GPS 受信機の方角をパラメータとして、誤差関数 E を勾配法によって最小化する。ただし、この誤差関数 E には、局所解が多数存在するため、単純な勾配法では、大域最小解を得ることが難しい。そこで、最適化の各フレームの重み μ_i を変化させることで、局所解を回避する。具体的には、GPS 測位値があるフレームの重みを大きく設定することおよび、すべてのフレームの重みを等価に設定することを、最適化処理が収束するステップごとに交互に繰り返す。

この局所解回避策は経験的に得たものであり、効果については予めシミュレーションによる予備実験により確認している。図 13 は、真値に誤差を加えることで人為的に作成した初期値を提案手法により最適化した際のカメラ位置の推定誤差の変化を表す。このシミュレーションは、3.4.1 節で示すシミュレーションと同様の条件で行い、重み μ_i は、500 ステップ毎に変化させている。同図から、重み変動を行わない場合は、カメラ位置の平均推定誤差が 200mm 程度のときに局所解に陥っているが、重み変動を行うことにより、局所解が回避されていることが分かる。

また、重み係数 ω を決定するために、予め重み係数 ω とカメラ位置の推定誤差の関係を予備実験により調べておく。予備実験では、撮影環境のスケールおよび特徴点の検出精度、GPS の測位精度は、後述する 3.4.1 節のシミュレーションと等しく設定し、外部パラメータの推定誤差を調べた。 ω の値を 10^{-7} から 10^{-15} まで調べた結果、 $\omega = 10^{-9}$ のときに最も真値に近いカメラ位置が得られた。この予備実験で使用した仮想環境と類似する環境で提案手法を用いる場合は、重み係数を $\omega = 10^{-9}$ に設定する。

3.3.2 各処理の詳細

本節では、先に定義した特徴点と GPS による位置誤差 E を用いたカメラパラメータの推定手順について、図 12 に示したフローチャートの各ステップ (A) ~ (D) を順に説明する。ステップ (A), (B) は毎フレーム、(C) は一定フレーム毎に実行され、(D) は最終的な最適化処理として 1 回のみ実行される。

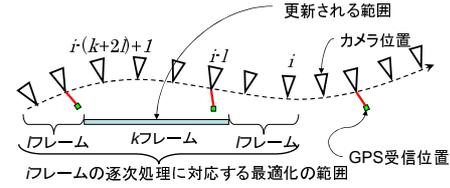


図 14: 最適化の範囲

(A) 特徴点の追跡

特徴点を入力画像上で検出し、前フレームと現フレームでの対応関係を求める。特徴点には、回転・拡大縮小等の画像の変形に対して頑健に同一の位置が検出されるエッジの角や交点を用いる。さらに、特徴点の仮対応づけの結果から統計的手法によって誤対応を排除し、それにより求まる暫定的な外部パラメータを用いて、特徴点の対応関係を修正する。これにより、特徴点の周辺類似パターンとの誤対応を防止する。具体的な処理手順としては、以下のとおりである。

1. Harris オペレータ [19] で求められる特徴量が極大値となる位置を特徴点の候補位置とする。
2. 前フレーム上の各特徴点の候補位置の近傍領域を現フレーム上でテンプレートマッチングによって探索し、現フレームにおける仮の追跡位置を決定する。本ステップにおける特徴点の追跡は、それぞれの単眼カメラの画像内でのみ行われる。
3. 特徴点の仮の追跡位置から、LMeds 基準を用いたロバスト推定 [20] によって、暫定外部パラメータを推定する。
4. 暫定外部パラメータを用いて、前フレームにおけるすべての特徴点の三次元位置から現フレーム上での出現位置を予測し、ステップ 2 よりも小さな探索窓を用いて対応づけを行い、追跡位置を修正する。ただし、ステップ 2 の追跡処理において画像からフレームアウトし追跡が失敗した特徴点についても、他のカメラの画像上に特徴点の三次元位置を投影することにより、出現位置を予測し、カメラ間の追跡処理を行う。

(B) 外部パラメータの初期値推定

処理 (A) で求めた現フレーム (第 i フレーム) 上の特徴点に関する再投影誤差の二乗和を最小化することで特徴点の三次元位置、および現フレームの外部パラメータである回転 R_i および並進 t_i を推定する。再投影誤差の二乗和は、以下のように表される。

$$\sum_c \sum_j w_j \Phi_{ijc}^2 \quad (14)$$

この初期値は、全ての単眼カメラに写る自然特徴点を統合的に扱い、線形解法と非線形最適化の組み合わせ [15] により推定される。

(C) GPS 測位値を用いた狭区間最適化

処理 (C) では、処理 (A)、処理 (B) で得られるカメラ外部パラメータを初期値として、前節で定義した誤差関数 E を最小化することで、GPS 測位値を反映させた外部パラメータを算出する。さらに、この処理で算出されたパラメータを処理 (A) にフィードバックすることで、特徴点对応づけの精度向上を図る。

具体的には、図 14 に示すように、処理 (A) および (B) で処理の対象となる第 i フレームに対して、第 $i - (k + 2l) + 1$ フレームから第 i フレームの特徴点追跡結果および GPS 測位値を用いて狭区間での誤差関数 E の最小化を行い、第 $i - (k + l) + 1$ フレームから第 $i - l$ フレームの k フレーム分の外部パラメータのみを更新する。これにより、更新する区間に対して、前後のフレームにおける特徴点の追跡結果および、GPS 測位値を最適化に反映させる。さらに、一度に複数のフレーム (k フレーム) を更新し、 k フレーム間隔で本処理 (C) を行うことで計算量の増大を軽減する。ただし、多数の追跡済みの特徴点が、推定結果を更新する区間 k フレーム内に写るように、 l は数十フレームに設定する。また、特徴点追跡が外部パラメータの蓄積誤差により破綻する前に GPS 測位値を反映させる必要があるため、 k は数フレームに設定する。

(D) GPS 測位値を用いた広範囲最適化

処理 (D) では、処理 (A) ~ (C) を繰り返すことで得られるカメラ外部パラメータと特徴点の三次元位置を初期値として、処理 (C) よりも広い区間で、外部パラメータの最適化を行う。これにより、特徴点追跡と GPS 測位の両情報に含まれる局所的な誤差の平均化を図る。ただし、本研究では、処理 (C) において間欠的に絶対指標を用いた最適化処理を行うため、時間的に十分離れ

た2フレームに関する外部パラメータや特徴点の三次元位置などの情報は、互いに影響しないと考えられる。そこで処理(C)での最適化範囲 $k+2l$ よりも十分に長い複数の $k'+2l'$ フレームのシーケンスに動画像を分割し、各シーケンスを独立に最適化する。ただし、処理(C)と同様に、最適化区間に外部パラメータを更新する区間 k' とその前後に更新しない区間 l' を設け、区間 l' を十分大きくとることで、各シーケンスの最適化における独立性を確保する。ただし、十分な数のGPS測位値を用いて測位誤差を平均化するために、 l' は数百フレームに設定する。また、分割されたシーケンス間の独立性を確保するために、 k' についても数百フレームに設定する。

3.4 実験

まず、式(13)で定義した誤差関数 E の最小化による最適化処理(C)の定量評価として、シミュレーションによる実験を示す。次に、特徴点検出位置およびGPS測位値の誤差モデルや特徴点対応付けの精度を含め、提案手法の有効性を確認するために、実環境で取得した動画像とGPS測位値からの外部パラメータの推定結果を示す。最後に、GPS測位値を用いた広範囲最適化(D)の有効性を確認する。

3.4.1 シミュレーションによる定量評価

シミュレーションによる評価では、提案手法の処理(D)によって推定される最終的な外部パラメータの推定精度を定量的に評価する。また、従来手法であるランドマーク・特徴点追跡による全方位型マルチカメラシステムの外部パラメータ推定手法[15]と精度を比較することにより、提案手法の有効性を確認する。

本シミュレーションで用いる仮想環境中の特徴点には、事前に屋外環境を撮影した全方位動画像を入力として佐藤らの手法[15]によって推定された特徴点の三次元位置を真値として用いた。この仮想環境内でカメラを移動させ、900フレームの特徴点検出画像を取得した。ただし、仮想カメラの内部パラメータには、前章で述べた手法により推定したLadybugのパラメータを用いた。カメラ座標系におけるGPS受信機の位置は、後述する実環境下での実験と同じ(60,-150,250)(単位mm)と設定し、GPS測位値には、上記GPS受信機の位置に誤差を付加し、これを1Hzに相当するフレーム間隔として15フレーム毎に与えた。特徴点の画像上での検出座標 q_{ij} および、GPS測位値 g_i に付加した誤差を表1に示す。

また、外部パラメータ R_i , t_i の初期値は、表2に示す誤差を真値に付加して作成した。これらの特徴点の検出誤差、カメラの位置および光軸の向きについては、佐藤らの屋外環境での実験結果[15]に基づき決定した。3.3.1節で定義した式(13)の誤差関数 E 中

の重み係数 ω は、前節で述べた値 10^{-9} とした。また、式(13)中の各フレームの重み係数 μ_i については、GPS測位値が得られていないフレームで1.0とし、GPS測位値が得られているフレームでは、2.0と1.0を交互に用いた。初期フレームにおける外部パラメータの初期値は、手動で大まかな姿勢を与え、位置に関してはGPS測位値を用いた。また、提案手法の精度評価のために、ランドマーク・特徴点追跡による手法[15]によって上述と同様の条件で外部パラメータを推定した。ただし、ランドマークは特徴点の信頼度 w_{ij} を十分大きくしたもののとして与え、全ての画像上での位置を与えた。また、ランドマークを与えるフレームの数や設定方法により外部パラメータの推定精度が変化する可能性があるため、全てのフレームで15個以上のランドマークが観測されるように設定した。以上の条件で、特徴点の三次元位置、外部パラメータ、カメラ座標系におけるGPS受信機の方位を最適化した。

図15および図16に初期値として与えた真値からの誤差、提案手法における誤差、ランドマーク・特徴点追跡による手法における誤差を示す。ただし、図15は、外部パラメータの位置に関する誤差、図16については、光軸の向きに関する誤差を示す。実験の結果、ランドマークを用いる手法では、カメラ位置に関しては平均誤差47.5mm、カメラ光軸の向きに関しては平均誤差0.0019radの精度で推定された。これに対して提案手法では、カメラ位置に関しては平均誤差30.7mm、カメラ光軸の向きに関しては平均誤差0.0023radの精度で推定された。

提案手法では、人手による計測が必要な事前知識を用いることなく、ランドマーク・特徴点追跡による手法と同程度の精度が得られた。両手法の推定精度の差の原因は、ランドマークを絶対指標とする手法とGPSを指標とする手法で、カメラの位置・姿勢の拘束のされ方が異なるためであると考えられる。本実験環境のようにランドマークが主に遠方に分布する場合、画像上でのランドマークの位置がカメラの並進の変化よりも回

表 1: 観測データに付加する誤差

特徴点の検出座標	標準偏差 1.6 画素のガウスノイズ + 量子化誤差
GPS の計測誤差	標準偏差 30mm のガウスノイズ

表 2: 外部パラメータの初期値に付加する誤差

カメラ位置	標準偏差 500mm のガウスノイズ
カメラ光軸の向き	標準偏差 0.020rad のガウスノイズ

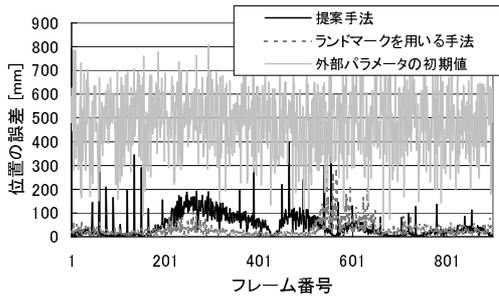


図 15: カメラ位置の推定誤差

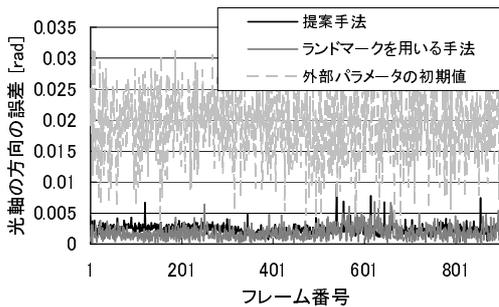


図 16: カメラ光軸方向の推定誤差

転の変化に敏感に反応することから、ランドマークを用いる場合は、姿勢に関する拘束が強く、結果として推定精度が高くなると考えられる。これに対して、GPSを用いる場合には、GPS 受信機の観測位置とカメラの投影中心間の距離が強く拘束されるため、カメラ位置に関する推定精度が従来手法よりも若干高まった反面、カメラ姿勢に関する精度は劣っていると考えられる。

3.4.2 実環境下での実験

本節では、提案手法における特徴点検出位置および GPS 測位値の誤差モデルや特徴点对応付けの精度を含め、提案手法の有効性を確認するために、実環境で取得した動画画像と GPS 測位値から外部パラメータを推定する。また、推定された外部パラメータを用いて動画画像上に仮想物体を重畳表示するマッチムーブを行うことで、外部パラメータが正しく推定されていることを確認する。

実験では、Ladybug と GPS 受信機 (Nikon LogPakII, 水平測位精度 $\pm 3.0\text{cm}$, 垂直測位精度 $\pm 4.0\text{cm}$) を図 18 に示すように電動スクータ (Segway LOC Segway) に固定して取り付けた。このような撮影機器を用いて、約 1.0km の距離を平均時速約 7.6km の速度で走行し、図 17 に示すような画像 7800 枚および、RTK 法により 1 秒間隔で計測された GPS 測位値を入力として用いた。ただし、動画画像と GPS 測位値は手動で開始時刻を同期させ、カメラと GPS 受信機間の距離は、手動で計測し



図 17: 取得した動画画像の 1 フレーム



図 18: 全方位型マルチカメラシステムと GPS 受信機

た 300mm を既知な情報として設定した。初期フレームにおける外部パラメータは、前節と同様に、姿勢に関しては大まかな初期値を与え、位置に関しては GPS 測位値を用いた。最適化処理 (C) においては、更新対象フレーム $k = 5$ および重複フレーム $l = 22$ に設定した。

まず、GPS 測位値を用いた最適化で得られた外部パラメータを初期値として動画画像全体で最適化した。全体最適化によって推定された特徴点の三次元位置と外部パラメータを図 19 に示す。

図中の曲線は推定されたカメラの移動経路を表し、点群は外部パラメータ推定に用いられた特徴点を表す。同図から分かるように、カメラパスは滑らかに推定されている。

次に、推定された外部パラメータを用いて仮想物体を代表カメラの入力動画画像に合成したマッチムーブの結果を図 20 に示す。仮想物体は、GPS による計測位置および GPS 受信機設置時の道路面からの高さから計算される道路面上に配置した。図 20 に示すシーンでは、仮想物体が現実環境中でほぼ一定の位置に合成されて

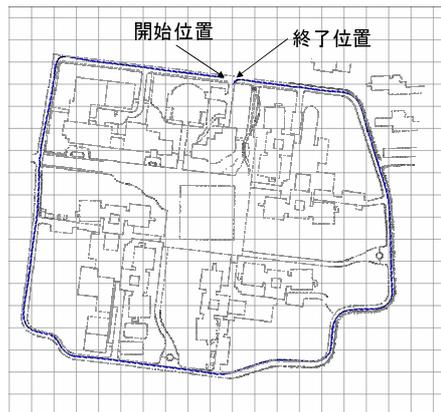


図 19: カメラ外部パラメータの推定結果

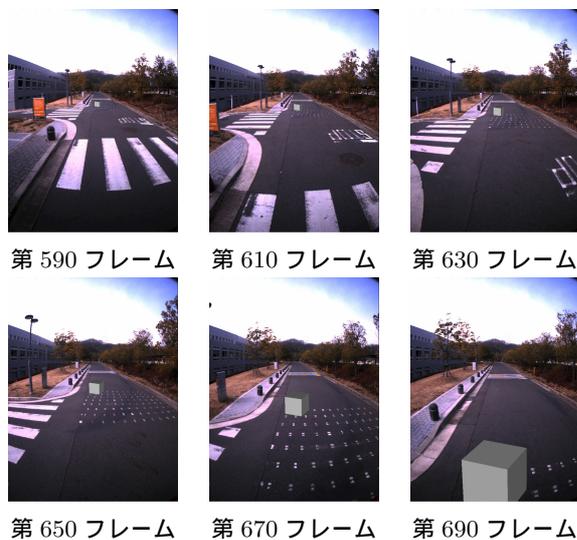


図 20: 推定された外部パラメータを用いたマッチングの結果

いることから、外部パラメータに大きな推定誤差が含まれないことを確認した。

最後に、処理 (D) における分割最適化の効果を検証するために、重複フレーム l' を 200 フレーム、更新フレーム k' を 1950 フレームとし、動画を 4 つのシーケンスに分割し、各シーケンスの統合結果を、動画を一括で最適化して得られる外部パラメータと比較した。図 21 に、分割最適化によって得られた外部パラメータの誤差を示す。本実験における平均誤差は、6mm であった。この値は、3.4.1 節で示した実験において示した、カメラ位置の推定誤差 32.9mm に対して小さく、分割最適化により一括での最適化に近い推定結果が得られていることが分かる。計算時間は、Pentium4 3GHz、メモリ 2GB の PC を用いた場合、一括最適化が約 14 時間であったのに対して、分割最適化では各シーケンスがそれぞれ約 3 時間であった。以上により、計算機に搭載されるメモリの量の制限など、一括で動画全体を最

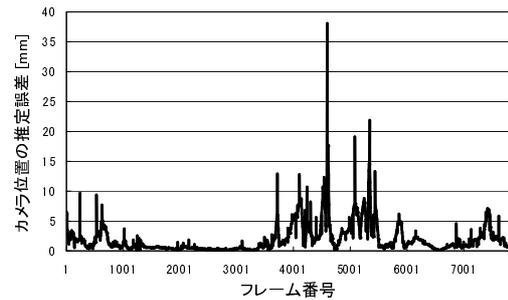


図 21: 分割最適化の精度

適化できないような場合においても、最適化の際に外部パラメータを更新しない区間 l' のフレーム数を十分大きく設定することで、一括で最適化する場合に近い結果が得られることが分かった。

3.5 結言

動画と GPS 測位値から全方位型マルチカメラシステムの外部パラメータを推定する手法を提案した。提案手法では、画像上で定義される特徴点の再投影誤差と GPS 測位値によって定義される位置誤差を同時に最小化することで、外部パラメータを各フレームに対して推定することができる。

シミュレーション実験および実環境データを用いた実験により、提案手法は、ランドマーク・特徴点追跡による手法に近い精度で外部パラメータを推定可能であることを確認した。提案手法により、現実環境に関する事前知識の取得に要する人的コストの大幅な削減が可能である。しかし、GPS の電波をさえぎる遮蔽物に囲まれるような状況下では、外部パラメータの推定精度の低下が考えられる。本研究では、遮蔽物などの影響を受ける場合については考慮していないが、このような大域的な誤差に対応するためには、GPS 測位値の信頼度等のセンサ情報を考慮したに誤差関数を導入する必要がある。

4 歩行装置を用いた能動的視点移動型テレプレゼンスシステム

4.1 緒言

本章では、広域実環境を仮想化し、1 つの経路上を移動する感覚を再現する能動的視点移動型テレプレゼンスシステムについて述べる。提案システムでは、高い臨場感を再現するための方策として、以下の 4 つの特徴を持つ。



図 22: 画像提示システムの外観

- (A) 広視野・高解像画像の提示 全方位型マルチカメラシステムにより取得し、生成した高解像度、広視野角な画像を提示する。
- (B) 提示画像のスタビライジング 歩行動作による視界変化の再現性の向上のために、歩行動作と関係の無い画像取得時における全方位型マルチカメラシステムの移動速度・姿勢の変動の除去する。
- (C) 歩行動作インタフェースの利用 歩行装置を用いた画像提示システムにより遠隔地での歩行動作中の視点位置・視線方向の変化による視界の変化の再現する。
- (D) ヘッドトラッキング 画像提示システム内での歩行動作中による頭部位置の変動を考慮した画像提示する。

本章では、まず、4.2節でシステムの実装法について説明し、4.3章では、試作システムの有効性を検証する実験を示す。最後に4.4節で本章をまとめる。

4.2 歩行装置を用いた画像提示システム

本研究で使用する画像提示システムは、図 23 に示すように、(a) 歩行動作インタフェース、(b) 描画用クライアント、(c) 没入型ディスプレイにより構成される。歩行動作インタフェースは、利用者の歩行動作を検出し、歩行装置を制御することで歩行運動によるユーザの移動を相殺する。描画用クライアント (CPU: Intel Pentium4 1.8GHz, Graphics Card: Geforce4 Ti4600) は、歩行動作インタフェースから受け取った情報を基にユーザの視点に応じた画像を没入型ディスプレイに描画する。以下に各部分の詳細を述べる。

(a) 歩行動作インタフェース

図 23(a) に示すように、歩行動作インタフェースは歩行装置 (限界速度約 1.6m/sec)、レーザレンジファイン

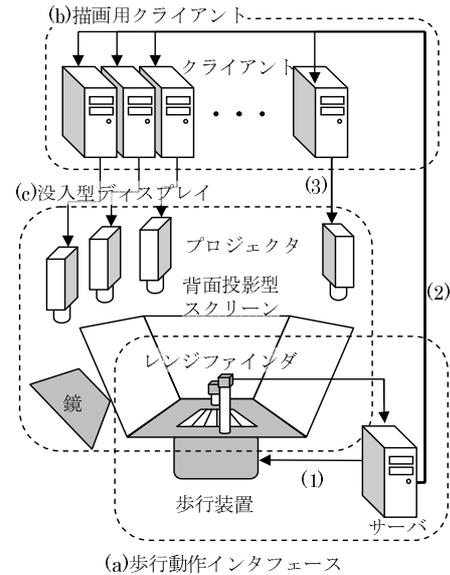


図 23: 歩行装置を用いた画像提示システム。

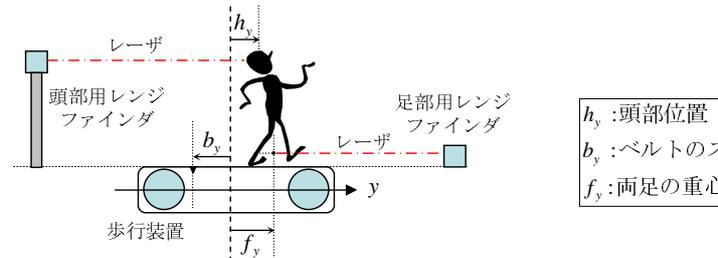


図 24: 歩行装置の座標系。

ダ (Sick LMS200) 2 台、サーバ (CPU: Intel Pentium4 2.4GHz) により構成される。サーバでは、レーザレンジファインダにより得られるレンジデータから図 24 に示すユーザの両足の重心位置 f_y 、頭部位置 h_y が算出され、これらに基づき歩行装置の制御 (図 23 中 (1)) および描画用クライアントの制御 (図 23 中 (2)) が行われる。以下では、まず、歩行動作の検出について述べ、次に歩行装置の制御法について述べる。

歩行動作は、図 24 に示すように、ユーザが立つ位置の前方および後方に設置されたレーザレンジファインダにより検出される。前方のものにより両足首、後方のものにより頭部が検出される。本研究で使用するレーザレンジファインダ LMS200 は、75Hz で 180 度の範囲 401 点の奥行きを計測できる。図 25 は足部の検出結果を示している。ただし、両図中の点はレンジデータを表し、長方形は設定した検出領域、放物線は曲線当てはめにより得た 2 次関数、横線は最終的な結果として使用する足部の検出位置を表す。LMS200 により取得したレンジデータから、 α, β, γ を係数とする 2 次関数 $y = \alpha(x + \beta)^2 + \gamma$ の曲線を当てはめ、頂点座標 γ を足

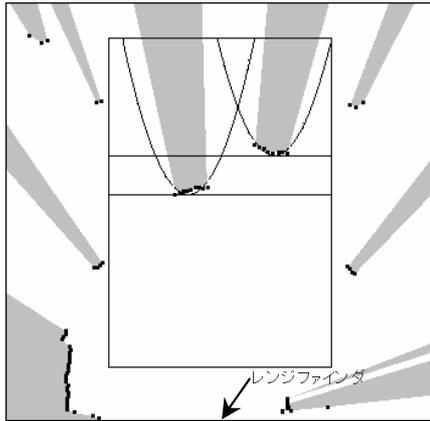


図 25: 足部のレンジデータと検出結果の例

部および頭部の位置として算出する。

歩行装置のベルトは利用者の両足の重心位置 f_y がベルト領域の中央に位置するように制御される [21]。ベルトのスライド幅が b_y ，仮想環境内でのユーザの視点移動量 v は $v = h_y - b_y$ で表される。サーバによって算出されたユーザの視点移動量 v は，全方位画像撮影時の全方位型マルチカメラシステムの移動量に対応し，逐次，フレームインデックスに変換され，頭部位置 h_y と共に描画用クライアントに送信される。

(b) 描画用クライアント

図 23 (b) に示すように，描画用クライアントは，12 台の PC により構成される。クライアントとサーバの間は，100Mbps の LAN により接続され，ネットワークを介してクライアントが制御される。頭部位置 h_y が 0 のときの画像を JPEG 形式で圧縮したものが，予めクライアントのハードディスクに蓄積されており，ユーザの頭部位置が変化したときは，グラフィックハードウェアの機能を利用して，頭部位置に応じてワーピングされた画像が描画される。

(c) 没入型ディスプレイ

図 23 (c) に示すように，没入型ディスプレイは高解像度な画像を提示するために，3 つの傾斜型平面スクリーンと 12 台のプロジェクタにより構成される。4 つのプロジェクタからの映像が反射鏡を介して 1 つのスクリーンに背面投影され，高解像度な画像がユーザに提示される。ユーザに広い視野を提供するために，スクリーンはユーザの前面および左右に配置されている。各プロジェクタの解像度は $1,024 \times 768$ である。ただし，複数のプロジェクタから 1 つのスクリーンに重複部分を持って画像が投影されるため，各スクリーンの解像度は，実質 200 万画素程度である。

4.3 カメラパラメータに基づく広視野角動画コンテンツの生成

提示画像の生成およびユーザ視点移動量と動画のフレームインデックスとの対応関係の算出について述べる。提示画像の生成は，基本的には第 2 章 2.4 節で述べた手法を用いるが，次に 3 点が異なる。

- スクリーンの形状を投影曲面とする
- 投影中心を頭部位置に合わせて動的に変更する
- 推定された全方位型マルチカメラシステムの姿勢を用いて，傾きや揺れによる効果を除去する

スクリーン内でのユーザの頭部位置を $[h_x, h_y, h_z, 1]^T$ ，スクリーン上の点を $P = [p_x, p_y, p_z, 1]^T$ ，点 P に対応する画像上の位置を u とすると，頭部位置 H から P への相対方向 $P' = [p_x - h_x, p_y - h_y, p_z - h_z, 0]$ を用いて，以下の式で表される。

$$u = d_c(N_{ci}R_iP') \quad (15)$$

ただし， $d_c(X)$ は，カメラ c のカメラ座標系での三次元位置 X から入力画像上での位置 u への変換関数である。 R_i は，没入型ディスプレイと撮影環境にそれぞれ固定された座標系との間の回転を表し，各フレーム i での全方位型マルチカメラシステムの移動経路に対する接線方向の画像がユーザの正面に提示されるように与える。

全方位型マルチカメラシステムの位置情報は，初期フレームからの全方位型マルチカメラシステムの移動量を算出し，ユーザの視点移動量 v とフレームインデックス i との対応関係 $i(v)$ を得るのに使用される。この対応関係 $i(v)$ は以下の手順で算出される。

1. 撮影時の Ladybug の投影中心の重心位置を各フレームごとに算出する
2. 重心位置の軌跡を平均フィルタでスムージングする
3. 各フレームに対する初期位置からの Ladybug の撮影経路に沿った距離を算出する

4.4 実験

実験では，主観評価実験を通して，4.1 節で述べたユーザの臨場感の向上のための方策である 4 項目 (A) ~ (D) それぞれの有効性を確認する。

実験に用いる提示画像は，3.4.2 節の実験で取得し，4.3 節で述べた手法により生成したものである。

4.4.1 主観評価による有効性の検証

構築した能動的視点移動型テレプレゼンスシステムでは、ユーザの視点移動が撮影系路上に制限されるため違和感が発生すると考えられる。この違和感により、4.1 節で述べたユーザの臨場感の向上のための方策である4項目 (A) 広視野・高解像画像の提示, (B) 提示画像のスタビライジング, (C) 歩行動作インタフェースの利用, (D) ヘッドトラッキングのそれぞれが無効化する可能性がある。本実験では、ユーザの臨場感を主観評価し、これらの項目の有効性を確認する。

評価方法

まず、試作システムの評価方法について説明する。テレプレゼンスシステム使用時の臨場感を絶対的な指標により定量化することは難しく、主観評価においても被験者の負担が大きい。そこで、2つのシステム X, Y を比較し、被験者はどちらのシステムを使用した時が臨場感が高かったかを「Xの臨場感が高い」「Yの臨場感が高い」「どちらとも言えない」の三択で評価する方式を採用する。被験者に説明する臨場感の定義は、「映像に写っている場所にいる感覚」とした。具体的には、比較用システム A~D を設定し、それぞれを試作システムと比較することにより、ユーザの臨場感の向上のための方策である4項目 (A) 広視野・高解像画像の提示, (B) 提示画像のスタビライジング, (C) 歩行動作インタフェースの利用, (D) ヘッドトラッキングの使用が、それぞれユーザの回答に影響するか調べる。比較用システムは、以下のように設定し、提案システムとしては、ユーザの臨場感の向上のための方策 (A)~(D) を採用したものをを用いた。

システム A 方策 (A) は採用せず、一般的なカメラの視野角および解像度を再現する意味で、640 × 480 画素相当の解像度画像を前面スクリーンのみに表示する。方策 (B)~(D) は採用する。

システム B 方策 (B) は採用せず、Ladybug の外部パラメータを全てのフレームで一定値に固定し、撮影中のカメラの姿勢変動等が提示画像に現れるようにする。方策 (A) および (C), (D) は採用する。

システム C 方策 (C) は採用せず、歩行装置のベルトを停止させ、市販のゲームコントローラにより歩行速度を 0~2.0[km/h] の範囲で指定できるようにする。方策 (A) および (B), (D) は採用する。

システム D 方策 (D) は採用せず、頭部位置が没入型ディスプレイの中央に固定されているものとして画像を提示する。方策 (A)~(C) は採用する。

ただし、これらの実験では、実験順序による結果の依存性を軽減するために、被験者は、事前に5分程度各

システムを体験し、各システムの評価実験間の切り替えを10秒以内にした。このように事前にシステムを体験させることで、被験者が評価実験中に学習する効果を軽減でき、評価実験の切り替えを短時間で実施することで、直前のシステムの使用感覚を忘却する効果を軽減できると考えられる。

評価結果

実験の目的および条件の詳しい知識を有しない10名を被験者として、上述の実験を行なった。評価結果を表3に示す。比較用システム A~C と試作システムの比較に関しては、有意水準5%のもとで統計的有意な差が確認された。これは、多くの被験者が試作システムの特徴である、(A) 広視野・高解像画像の提示, (B) 提示画像のスタビライジング, (C) 歩行動作インタフェースの利用が、臨場感の向上という点について有効であることを示している。ただし、システム D との比較に関しては、「試作システムの臨場感が高い」と答えた被験者の数が「比較用システムの臨場感が高い」と答えた被験者の数と同じであり、「どちらとも言えない」と答えた被験者が一番多かった。これは、比較用システム D には採用されていない (D) ヘッドトラッキングの機能が、試作システム上での効果が低かったことを意味する。

ヘッドトラッキングの効果が低い点に対する考察を述べる。この結果の主な原因は、歩行中に被験者の頭部位置があまり動かなかったことが考えられる。実際、被験者の頭部位置は、一定速度で歩いている時は、標準偏差10cm程度の変動であり、頭部位置からスクリーンまでの距離の10%以下であった。逆に、頭部位置を最大幅150cm、標準偏差32cmで動かす1人の被験者は、「試作システムの臨場感が高い」に投票している。このことから遠隔地での散歩を再現するシステムやトレーニング支援など、特に頭部位置を積極的に動かさないアプリケーションでは、ヘッドトラッキングの効果は小さいと考えられる。

4.5 結言

本章では、臨場感の高い能動的視点移動型テレプレゼンスシステムの構築を目的とし、全方位型マルチカメラシステムにより取得した画像を用いたシステムについて述べた。実験では、主観評価実験により、(A) 広視野・高解像画像の提示, (B) スタビライジング, (C) 歩行動作インタフェースの利用, (D) ヘッドトラッキング

表 3: 評価結果

比較用システム	A	B	C	D
比較用システムの臨場感が高い	0	0	2	1
どちらとも言えない	1	0	0	8
試作システムの臨場感が高い	9	10	8	1

のそれぞれを利用する効果を確認した。その結果、少なくとも (A) ~ (C) の利用により、多くの被験者の臨場感が向上することを確認した。このことは、能動的視点移動型テレプレゼンスシステムの1つである歩行装置を用いたシステムにおいて、前章までに提案するカメラ内部・外部パラメータの推定技術を利用することが臨場感の向上という点に対して有効であることを示している。

5 結論

本研究では、遠隔地の広域実環境下において単一経路上での移動感覚を再現するテレプレゼンスシステムを対象として、全方位型マルチカメラシステムを導入し、ユーザの臨場感の向上に関係する技術の開発を行った。本報告では、テレプレゼンスシステムをユーザが受動的に視点移動する受動的視点移動型テレプレゼンスシステムと能動的に視点移動する能動的視点移動型テレプレゼンスシステムの2つの方式に区別し、両方式に共通する技術として、全方位型マルチカメラシステムの内部パラメータの推定について述べ、特に能動的視点移動型テレプレゼンスシステムに対しては、画像取得時の全方位型マルチカメラシステムのカメラ外部パラメータの推定について述べた。実際に受動的および能動的視点移動型の両方式のテレプレゼンスシステムを構築し、上述の技術により臨場感の高いテレプレゼンスシステムが実現可能であることを示した。これらを利用することにより、従来提案されている様々なテレプレゼンスシステムにおける臨場感の向上が期待できる。今後は、複数経路の統合による視点移動範囲の拡大や動物体の表現について検討する。

参考文献

- [1] T. Kawanishi, K. Yamazawa, H. Takemura and N. Yokoya: "Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors," Proc. 14th Int. Conf. on Pattern Recognition (ICPR '98), 1, pp. 445-489, 1998.
- [2] 戎野 聡一, 山澤 一誠, 竹村 治雄, 横矢 直和: "円筒パターンを用いた全周ステレオ画像センサのキャリブレーション手法", 日本バーチャルリアリティ学会第4回大会論文集, pp. 211-212, 1999.
- [3] 棚橋 英樹, 佐藤 淳, 王 彩華, 丹羽 義典, 山本和彦: "全方向ステレオシステム (SOS) のキャリブレーション手法", 映像メディア学会誌, 56, 4, pp. 603-610, 2002.
- [4] S. Güven and S. Feiner: "Authoring 3D hypermedia for wearable augmented and virtual reality," Proc. 7th IEEE Int. Symp. on Wearable Computers, pp. 118-126, 2003.
- [5] A. J. Davison: "Real-time simultaneous localisation and mapping with a single camera," Proc. 9th IEEE Int. Conf. on Computer Vision (ICCV2003), 2, pp. 1403-1410, 2003.

- [6] T. Sato, M. Kanbara, N. Yokoya and H. Take-mura: "Dense 3-D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera," Int. Jour. of Computer Vision, 47, 1-3, pp. 119-129, 2002.
- [7] A. I. Comport, É. Marchand and F. Chaumette: "A real-time tracker for markerless augmented reality," Proc. 2nd ACM/IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR2003), pp. 36-45, 2003.
- [8] L. Vacchetti, V. Lepetit and P. Fua: "Combining edge and texture information for real-time accurate 3D camera tracking," Proc. 3rd IEEE and ACM Int. Symp. on Mixed and Augmented Reality (ISMAR2004), pp. 48-57, 2004.
- [9] Point Grey Research, Inc.: Ladybug Omnidirectional Camera System User Guide version 1.0, 2002.
- [10] R. Y. Tsai: "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," IEEE Jour. of Robotics and Automation, RA-3, 4, pp. 323-344, 1987.
- [11] T. Sato, M. Kanbara, N. Yokoya and H. Take-mura: "Dense 3D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera," Int. Jour. of Computer Vision, 47, 1-3, pp. 110-129, 2002.
- [12] 出口 光一郎: "射影幾何学による PnP カメラ補正問題の統一的解法", コンピュータビジョン '90 シンポジウム 論文集, pp. 41-50, 1990.
- [13] B. K. P. Horn: Robot Vision, chapter 10, pp. 206-209, Mit Press, 1986.
- [14] N. Asada, A. Amano and M. Baba: "Photometric calibration of zoom lens systems," Proc. Int. Conf. Pattern Recognition, A, pp. 186-190, 1996.
- [15] 佐藤 智和, 池田 聖, 横矢 直和: "複数動画からの全方位型マルチカメラシステムの位置・姿勢パラメータの推定", 電子情報通信学会論文誌 (D-II), J88-D-II, 2, pp. 347-357, 2005.
- [16] A. W. Fitzgibbon and A. Zisserman: "Automatic camera recovery for closed or open image sequences," Proc. 5th European Conf. on Computer Vision, I, pp. 311 - 326, 1998.
- [17] M. Pollefeys, R. Koch, M. Vergauwen, B. Dekenuydt and L. V. Gool: "Three-dimensional scene reconstruction from images," Proc. SPIE, 3958, pp. 215-226, 2000.
- [18] B. Triggs, R. H. P. McLauchlan and A. Fitzgibbon: "Bundle adjustment a modern synthesis," Vision Algorithms: Theory and Practice, pp. 298 - 375, 2000.
- [19] C. Harris and M. Stephens: "A combined corner and edge detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- [20] 栗田 多喜夫, 坂上 勝彦: "ロバスト統計とその画像理解への応用", 画像の認識・理解シンポジウム (MIRU2000) 講演論文集, Vol. I, pp. 65-70, 2000.
- [21] H. Iwata: "Walking about virtual environments on an infinite floor," Proc. IEEE Virtual Reality '99, pp. 286-293, 1999.