

# GPS 測位精度を考慮した動画像と GPS の併用による カメラ位置・姿勢推定

条 秀行<sup>†</sup> 武富 貴史<sup>†</sup> 佐藤 智和<sup>†</sup> 横矢 直和<sup>†</sup>

<sup>†</sup> 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: †{hideyuki-k,takafumi-t,tomoka-s,yokoya}@is.naist.jp

あらまし 本稿では、広域屋外環境における動画像からのカメラ位置・姿勢推定において、GPS を併用することで蓄積誤差を抑止する手法を提案する。従来提案されている動画像と GPS の併用手法は、GPS の測位誤差として平均値 0 の正規分布を仮定しているため、特に GPS の測位精度が低い場合などのように、短時間の観測において平均測位位置が真値から大きく外れる (バイアス誤差が大きい) 場合においてカメラ位置・姿勢の推定精度が大きく低下するという問題がある。この問題に対して、本研究では、GPS 受信機の真の位置が GPS の測位位置を中心とする GPS の測位精度に依存した一定の範囲内に存在することを仮定し、この仮定に基づくエネルギー関数を最小化することで、バイアス誤差の影響を排除しながら GPS の測位精度に応じて動画像からのカメラの位置・姿勢推定結果を補正する。  
キーワード カメラ位置・姿勢推定, structure from motion, GPS

## Extrinsic Camera Parameter Estimation Using Image Sequence and GPS Considering GPS Positioning Accuracy

Hideyuki KUME<sup>†</sup>, Takafumi TAKETOMI<sup>†</sup>, Tomokazu SATO<sup>†</sup>, and Naokazu YOKOYA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, 630-0192, Japan

E-mail: †{hideyuki-k,takafumi-t,tomoka-s,yokoya}@is.naist.jp

**Abstract** This paper proposes a method for estimating extrinsic camera parameters using image sequence and position data acquired by GPS. In conventional methods, the accuracy of the estimated camera position largely depends on the accuracy of GPS positioning data because they assume that GPS position error is very small or normally distributed. In order to solve this problem, in this study, we have employed a simple assumption that true GPS position exists within a certain range from the observed GPS position and the size of the range depends on the GPS positioning accuracy. Concretely, the proposed method estimates camera parameters by minimizing an energy function that is defined by using the reprojection error and the penalty term for GPS positioning.

**Key words** extrinsic camera parameter estimation, structure from motion, GPS

### 1. はじめに

現実環境を撮影した動画像に対するカメラの位置・姿勢情報は三次元形状復元や自由視点画像生成、拡張現実感をはじめとする多くの分野への応用が可能である。近年これらの分野では、広域な屋外環境への手法の適用が試みられ [1-3]、広域な屋外環境を移動しながら撮影された動画像に対するカメラの位置・姿勢を高精度に推定することが求められている。しかし、撮影された動画像のみからカメラの位置・姿勢を推定する Structure from Motion 法 [4,5] や Visual SLAM 法 [6,7] には、推定されるカメラ位置・姿勢に誤差が蓄積するため、特に広範囲を移動する場合に推定精度が低下するという問題がある。

このような問題に対し、動画像に加えて環境に関する事前知識や GPS を併用する手法 [1,8-13] が提案されている。環境に関する事前知識を併用する手法 [8,9] では、三次元位置が既知の基準点 [8] やシーン中の一部の領域に対するワイヤフレームモデル [9] を利用する。これらの手法は、他のセンサなどを必要とせず、カメラの絶対位置・姿勢を推定できるが、広範囲な屋外環境での利用においては基準点やモデルの三次元位置を計測するために多くの人的コストがかかるという問題がある。一方、GPS を併用する手法 [1,10-13] は、動画像中の特徴点の追跡によって得られるカメラの位置・姿勢と GPS から得られる位置情報を同時に用いることで、環境に対する事前知識なしに絶対的なカメラ位置・姿勢を推定するこ

とができる。従って、広域屋外環境でのカメラ位置・姿勢の推定には、動画像と GPS を併用する手法が適している。これら動画像と GPS を併用する手法は、情報の融合にカルマンフィルタを用いる手法 [1, 10, 11] とバンドル調整 (bundle adjustment) を用いる手法 [12, 13] に大別できる。

カルマンフィルタを用いる手法 [1, 10, 11] は動画像と GPS の情報を高速に融合することができるため、拡張現実感などのリアルタイムアプリケーションに適している。Agrawal ら [10] は、動画像から推定されるカメラ位置と 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。Pollefeys ら [1] は、拡張カルマンフィルタを用いた動画像からの推定手法を基礎とし、GPS、慣性航法装置からの観測をカルマンフィルタに組み込むことにより、より密な情報の融合を実現している。これらカルマンフィルタに基づく手法は、過去の状態と現在の観測から現在の状態を逐次的に推定する。このため、入力全体で最適化を行い精度を高めることは難しい。

一方、バンドル調整を用いる手法 [12, 13] は、動画像からのカメラ位置・姿勢推定において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することで、動画像と GPS の情報を融合する。横地ら [12] は、GPS の測位誤差が単一の正規分布で近似できることを前提に、定義した誤差関数を最小化することで、カメラ位置・姿勢を推定している。この手法は、動画像全体でカメラ位置・姿勢を最適化することができる。しかし、実際には、GPS の測位精度は GPS 電波の遮蔽や反射、衛星の配置等によって大きく変化するため、単一の正規分布により測位誤差をモデル化することは難しく、特に GPS の測位精度が低い場合にカメラ位置・姿勢の推定精度が大きく低下するという問題がある。これに対して、Anai ら [13] は GPS 受信機から得られる精度情報に応じて正規分布の分散を変化させることにより、測位精度の変化を考慮する手法を提案している。しかし、短時間の観測において測位に大きなバイアス誤差が含まれる場合において、推定精度が低下するという問題は依然として解消されていない。これは、GPS の測位誤差として平均値 0 の正規分布を仮定する一方で、実際の短時間における GPS の測位誤差の分布がこのような正規分布で適切に近似できないことに起因している。

本研究では、三次元形状復元や自由視点画像生成などオフラインで高精度なカメラ位置・姿勢を必要とするアプリケーションへの適用を考え、広範囲な環境を移動しながら撮影した動画像に対するカメラの位置・姿勢推定において、比較的高精度な絶対位置情報を推定可能なバンドル調整による GPS と動画像情報の融合手法 [12, 13] を基礎とし、GPS の測位精度が低い場合にカメラ位置・姿勢の推定精度が低下するという問題を解決する。具体的には、GPS の測位誤差を単一の正規分布でモデル化する代わりに、GPS 受信機の真の位置が GPS の測位位

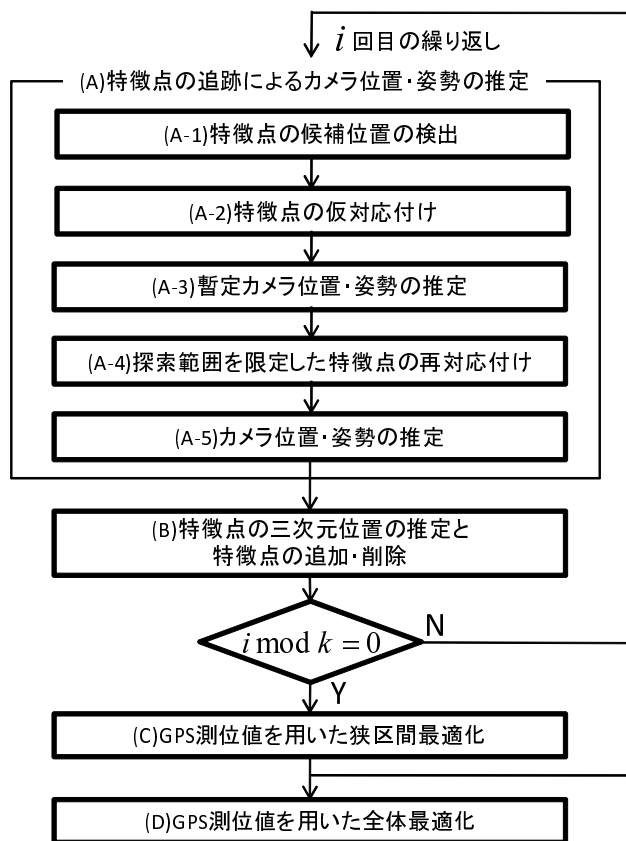


図 1 提案手法の処理の流れ

置を中心とする、GPS の測位精度 (RTK-fix, RTK-float 等) に依存した一定の範囲内に存在することを仮定し、この仮定に基づくエネルギー関数を最小化することで、GPS の測位精度に応じて動画像からのカメラの位置・姿勢推定結果を補正する。

以下、2 節では、GPS の測位精度を考慮したエネルギー関数の定義とカメラ位置・姿勢推定手法について詳述する。3 節では、屋外環境を移動しながら撮影した動画像を用いてカメラの位置・姿勢推定を行い、従来手法と比較することで精度を検証する。最後に 4 節でまとめと今後の課題について述べる。

## 2. GPS の測位精度を考慮したカメラ位置・姿勢推定

本研究では、カメラと GPS 受信機の相対位置関係を固定して撮影した動画像と GPS の測位値から、GPS 測地座標系におけるカメラの位置・姿勢を推定する。ただし、カメラと GPS 受信機の相対位置関係は既知であり、動画像と GPS の測位値は同期して得られるものとする。

提案手法はバンドル調整による GPS と動画像情報の融合手法 [12, 13] を基礎とし、これらの手法と同様に、特徴点の再投影誤差と GPS の測位位置に関するエネルギーを同時に最小化することにより、動画像からのカメラ位置・姿勢の推定結果を補正する。提案手法の処理の流れを図 1 に示す。まず、フレーム毎の処理として、特徴点の追跡によるカメラ位置・姿勢推定 (A)、特徴点の

三次元位置の推定と特徴点の追加・削除 (B) を行う。ここで、GPS の測位値が得られる一定フレーム間隔  $k$  毎に、GPS の測位値を用いた狭区間最適化 (C) を行い、カメラの位置・姿勢と特徴点の三次元位置を最適化する。これにより、逐次的に蓄積誤差を抑えたカメラの位置・姿勢を得る。これらの処理 (A) ~ (C) を初期フレームから最終フレームまで繰り返すことで、逐次的にカメラ位置・姿勢および特徴点の三次元位置の初期値を得た上で、最後に動画全体での最適化処理を行う (D)。以下では、まず、本研究で新たに提案する GPS の測位精度を考慮したエネルギー関数について述べる。次に、処理 (A) ~ (D) について詳述する。

## 2.1 GPS の測位精度を考慮したエネルギー関数の定義

本研究では、動画を対象とした三次元復元の評価尺度として広く用いられている再投影誤差と、GPS に関して本研究で新たに定義するエネルギー項の和を最小化することで、カメラ位置・姿勢および特徴点の三次元位置を推定する。具体的には、エネルギー関数  $E$  を、 $i$  フレームにおける再投影誤差  $\Phi_i$  と GPS 測位位置に関するペナルティ項  $\Psi_i$  を用いて以下のように定義する。

$$E = \sum_{i \in F} \Phi_i + \omega \sum_{i \in F_g} \Psi_i \quad (1)$$

ただし、 $\omega$  は重み係数、 $F$  は動画のフレームの集合、 $F_g$  は GPS 測位時刻に対応する動画のフレームの集合を表す。以下、再投影誤差に関するエネルギー  $\Phi_i$  と GPS 測位位置に関するペナルティ項  $\Psi_i$  について詳述する。

### 2.1.1 再投影誤差

再投影誤差は、特徴点の三次元位置を画像上に投影した座標と、画像上で検出された特徴点の座標の距離の二乗和であり、動画を対象としたバンドル調整において一般的に用いられている。本研究では、再投影誤差に関するエネルギー  $\Phi_i$  を以下のように定義する。

$$\Phi_i = \frac{1}{|P_i|} \sum_{j \in P_i} \mu_j (\mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij})^2 \quad (2)$$

ただし、 $P_i$  は第  $i$  フレームの画像上で検出される特徴点の集合である。また、 $\mathbf{q}_{ij}$  は、2.2.1 で述べる手法により検出された、第  $i$  フレームにおける特徴点  $j$  の画像上での座標、 $\hat{\mathbf{q}}_{ij}$  は特徴点  $j$  の三次元位置を画像上へ投影した座標を表す。 $\mu_j$  は逐次処理で得られた特徴点  $j$  の信頼度であり、特徴点  $j$  が第  $(w - m)$  フレームから第  $w$  フレームまで追跡された場合において、以下の式で算出される [8]。

$$\mu_j = \frac{1}{2\sigma_j^2} = \frac{m+1}{2} \left( \sum_{i=w-m}^w (\mathbf{q}_{ij} - \hat{\mathbf{q}}_{ij})^2 \right)^{-1} \quad (3)$$

ただし、 $\sigma_j^2$  は特徴点  $j$  の再投影誤差の分散を表す。

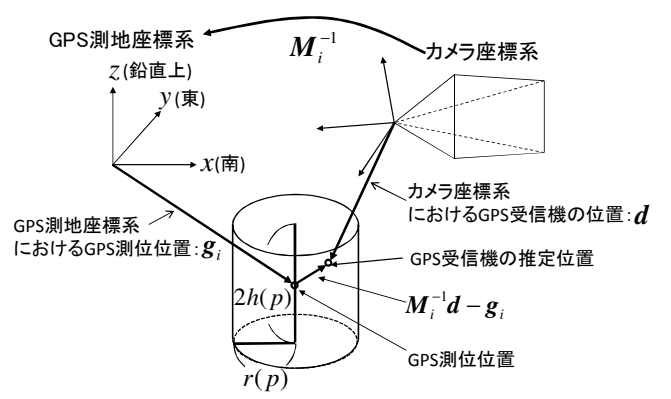


図2 GPS 測位位置に関するペナルティ

### 2.1.2 GPS 測位位置に関するペナルティ

従来手法 [12, 13] は、GPS の測位誤差が平均値 0 の正規分布で近似できることを前提に、以下のような GPS 測位位置に関するペナルティ項を定義している。

$$\hat{\Psi}_i = (M_i \mathbf{g}_i - \mathbf{d})^2 \quad [12] \quad (4)$$

$$\hat{\Psi}_i = \frac{1}{\tau^2} (M_i \mathbf{g}_i - \mathbf{d})^2 \quad [13] \quad (5)$$

ただし、 $M_i$  は第  $i$  フレームにおける GPS 測地座標系からカメラ座標系への変換行列、 $\mathbf{d}$  はカメラ座標系における GPS 受信機の位置、 $\mathbf{g}_i$  は第  $i$  フレームの GPS 測地座標系における GPS の測位位置を表す。また、 $\tau$  は GPS 受信機から得られる GPS 測位位置の RMS を表す。前述したように、これらのペナルティ項を用いた場合において、GPS の平均測位位置が真値から大きく外れる誤差が生じた場合に、推定精度が低下するという問題がある。

本研究では、GPS の測位誤差を正規分布でモデル化する代わりに、GPS 受信機の真の位置が GPS の測位位置を中心とする、GPS の測位精度 (RTK-fix, RTK-float 等) に依存した一定の範囲内に存在するという仮定を用いる。一般に GPS の測位誤差は水平方向と高度方向で異なることから、本稿では、GPS の水平方向の誤差と高度方向の誤差を独立に扱う。すなわち、図 2 に示すように、一定の範囲を円柱であるとし、GPS 測位位置に関するエネルギー  $\Psi_i$  を以下のように定義する。

$$\Psi_i = \left( \frac{1}{r(p)} \sqrt{x_i^2 + y_i^2} \right)^{2n} + \left( \frac{1}{h(p)} z_i \right)^{2n} \quad (6)$$

$$\begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = M_i^{-1} \mathbf{d} - \mathbf{g}_i \quad (7)$$

ただし、 $M_i^{-1}$  は第  $i$  フレームにおけるカメラ座標系から GPS 測地座標系への変換行列、 $r$  は円柱の半径、 $h$  は円柱の高さの 1/2、 $p$  は GPS の測位精度を表す。また、 $n$  はあらかじめ与える定数であり、 $n$  に大きな自然数を与えることで、GPS 受信機の位置が測位位置を中心と

した円柱の中に存在する場合には  $\Psi_i$  が 0 に近い値となり、円柱の外となる場合には非常に大きな値となる。ここでは、 $\Psi_i$  を微分可能な形で定義することで、後述する最適化処理における勾配計算を可能としている。

なお、GPS の精度に影響を与えるものとして、GPS の種類 (単独測位, DGPS, RTK-GPS), RTK-GPS における解の種類 (RTK-fix, RTK-float), DOP などが考えられる。後述する実験では、GPS として RTK-GPS を用いるが、この場合には解の種類に応じて測位精度が大きく異なることが知られているため、これを測位精度  $p$  として用いる。また、GPS の測位誤差はさまざまな要因によるため [14, 15], 理論的に大きさを見積もることは難しい。従って、それぞれの測位精度  $p$  における円柱のパラメータ  $r(p), h(p)$  は、カメラ位置・姿勢を推定する環境において GPS の定点観測を行うことで実験的に決定する必要がある。ただし、本研究では、GPS の測位値に含まれる、少数の誤差が非常に大きい外れ値は、GPS からの情報と動画像からの情報の整合性を検証することにより除去できるものとする。

## 2.2 動画像と GPS の併用によるカメラ位置・姿勢推定

本節では、図 1 に示した提案手法の処理の流れの各処理 (A)~(D) について詳述する。処理 (A), (B) はフレーム毎、処理 (C) は一定フレーム毎、処理 (D) は最終的な最適化処理として 1 回のみ実行される。

### 2.2.1 特徴点の追跡によるカメラ位置・姿勢の推定

逐次処理 (A) においては、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対応関係を求めることにより、カメラの位置・姿勢を推定する。本研究では特徴点として、回転・拡大などの変形に対してロバストに同一の点を検出可能な輝度エッジのコーナーを用いる。また、特徴点の仮の追跡を行い、ロバスト推定を用いて暫定的なカメラ位置・姿勢を推定することで、追跡すべき特徴点の探索範囲を限定し、周辺類似パターンとの誤対応を防止する [8]。以下に具体的な処理の流れを示す。

- (A-1) Harris オペレータ [16] の出力が極大となる点を特徴点の追跡位置の候補とする。
- (A-2) 前フレームにおける各特徴点の近傍領域を現フレーム上で SSD (Sum of Squared Differences) を評価尺度とした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。
- (A-3) 仮の追跡位置から、LMedS 基準を用いたロバスト推定 [17] によって誤対応を排除し、暫定的なカメラの位置・姿勢を得る。
- (A-4) 暫定的なカメラの位置・姿勢と特徴点の三次元位置から、現フレームにおける特徴点の出現位置

を予測し、周辺の小さな領域に対して対応付けを行い、追跡位置を確定する。

- (A-5) 追跡された特徴点の画像上の位置と三次元位置から、カメラの位置・姿勢を推定する。

(A-3), (A-5) におけるカメラ位置・姿勢は、追跡された特徴点の画像上の座標と特徴点の三次元位置から、以下の式を最小化することで推定する。

$$E_i = \sum_{j \in P_i} \mu_j (q_{ij} - \hat{q}_{ij})^2 \quad (8)$$

$E_i$  は、GPS の測位値を陽に用いることなく算出されるが、2.2.3 で述べる GPS 測位値を用いた最適化処理を一定フレーム毎に実行することで、カメラ位置・姿勢と特徴点の三次元位置が GPS 測地座標系上に配置され、最終的には GPS 測地座標系におけるカメラの位置・姿勢を求めることができる。

### 2.2.2 特徴点の三次元位置の推定と特徴点の追加・削除

現フレームまでのカメラ位置・姿勢と特徴点の画像上での検出座標を用いて、以下の式を最小化することにより、特徴点の三次元位置を推定し、更新する (処理 B)。

$$E_j = \sum_{i \in S_j} (q_{ij} - \hat{q}_{ij})^2 \quad (9)$$

ただし、 $S_j$  は特徴点  $j$  が検出されたフレームの集合を表す。

また、本研究では、広域な屋外環境を対象とした動画像を扱うため、カメラの動きによって特徴点の出現・消失が起こる。そこでフレーム毎に複数の評価尺度を用いて自動で特徴点の追加・削除を行う [8]。なお、本処理では処理 (A) と同様に、GPS 測位値を陽に使用していないが、GPS 測位値を用いた最適化によって、最終的には GPS 測地座標系における特徴点の三次元位置が得られる。

### 2.2.3 GPS 測位値を用いた最適化

GPS 測位値を用いた最適化には、逐次処理において範囲を限定して行う狭区間最適化 (C) と、逐次処理の終了後、動画像全体を対象として行う全体最適化 (D) がある。

処理 (A), (B) は、動画像上の情報のみを考慮した推定を行うため、得られたカメラの位置・姿勢と特徴点の三次元位置には蓄積誤差が存在する。従って、処理 (A), (B) を繰り返すだけでは、以下の問題が生じる。

- (1) 蓄積誤差の影響により、2.2.1 で述べた特徴点の追跡に失敗する。
- (2) 逐次処理の終了後に行う全体最適化 (D) において、初期値が真値から離れるため、大域最適解を求めることが難しくなる。

そこで処理 (C) では、逐次処理において、第  $(i-1)$  フレームから現フレーム (第  $i$  フレーム) を用いて、式 (1) で定義した誤差関数  $E$  を勾配法によって最小化することにより、GPS の測位値を用いてカメラの位置・姿勢と特徴点の三次元位置を最適化する。

処理 (D) では、処理 (A) ~ (C) で得られたカメラ位置・姿勢、特徴点の三次元位置、特徴点の信頼度を初期値として、動画像全体を用いて誤差関数  $E$  を最小化することにより、カメラ位置・姿勢の推定精度の向上を図る。

### 3. 実験

本節では、移動撮影により取得した動画像と RTK-GPS の測位値を用いてカメラ位置・姿勢の推定を行い、従来手法との定量的な推定精度の比較を行うことで、提案手法の有効性を示す。ここではまず、GPS を固定して長時間観測を行うことで、2.1.2 で述べた提案手法で用いる測位精度ごとの位置計測の誤差範囲を決定し、次に決定したパラメータを用いた推定結果により評価を行う。

#### 3.1 定点観測による GPS 測位値の誤差分析

本実験では、GPS 測位値の定点観測を行うことで、式 (6) で定義した GPS の測位位置に関するペナルティで用いる、測位精度ごとの円柱のパラメータ  $r(p), h(p)$  を決定する。ここでは、GPS の測位精度  $p$  として GPS の解の種類 (RTK-fix, RTK-float) を用いる。

##### 3.1.1 実験条件

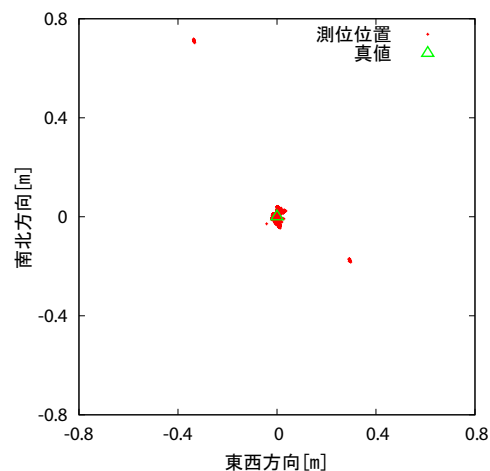
RTK-GPS 受信機 (TOPCON GR-3) を固定し、約 5 時間の連続測位を行った。表 1 に、利用した RTK-GPS 受信機のカタログ仕様を示す。測位は、後述する実験においてカメラ位置・姿勢推定を行う環境と概ね同じ環境で行った。GPS の測位周期は 1Hz であり、インターネット回線を通して補正データを受信することで、1 台の RTK-GPS 受信機で RTK 測位が可能な仮想基準点方式で測位を行った。

##### 3.1.2 GPS の測位精度に依存した誤差範囲の決定

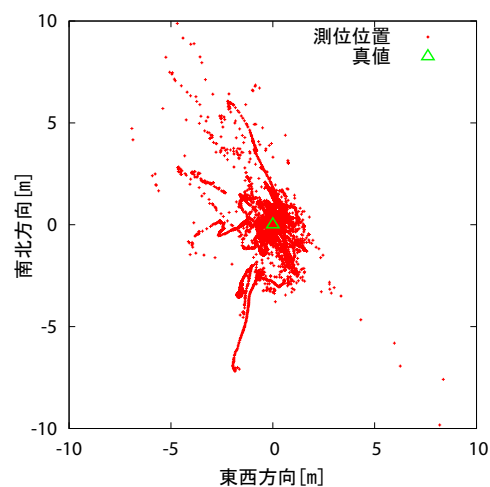
本実験では、得られた全ての RTK-fix データの平均位置を定点観測を行った地点の真値とし、各測位値に対する誤差を算出した。定点観測によって得られた GPS の測位位置の分布を、水平方向について測位精度ごとに図 3 に示す。また、表 2(a) に、測位精度ごとの真値からの誤差の最大値を示す。これらの結果から、測位精度によって誤差のスケールが大きく異なることが分かる。表

表 1 RTK-GPS 受信機のカタログ仕様

機種名	GR-3
使用搬送波	GPS L1/L2/L2C (2 周波) GLONASS L1/L2 (2 周波)
水平測位精度 (RTK)	$\pm(10 \text{ mm} + 1.0 \text{ ppm} \times D)$ m.s.e D:測定距離
垂直測位精度 (RTK)	$\pm(15 \text{ mm} + 1.0 \text{ ppm} \times D)$ m.s.e D:測定距離



(a) RTK-fix



(b) RTK-float

図 3 定点観測における水平方向の測位位置の分布

2(b) に、測位精度ごとに、水平方向、高度方向それぞれについて誤差が大きい 5% の測位値を除去した結果を示す。外れ値を除去することにより、誤差の最大値が大幅に小さくなっていることおよび、図 3 に示した測位位置の分布から、GPS の測位値には誤差が極端に大きい外れ値が含まれていることが確認できる。以降の実験では、GPS 測位の外れ値は動画像からの情報と GPS からの情報の整合性を検証することにより除去できるものとし、表 2(b) に示した、各測位精度における水平方向の誤差の最大値と高度方向の誤差の最大値を、それぞれ円柱のパラメータ  $r(p), h(p)$  とする。

ただし、以降の実験では、カメラと GPS 受信機は移動しているため、 $r(p), h(p)$  の決定にはこの移動量を考慮する必要がある。すなわち、本研究で用いる式 (1) のエ

表 2 測位精度ごとの誤差の最大値 [mm]

測位精度	(a) 全データ		(b) 外れ値除去後	
	水平方向	高度方向	水平方向	高度方向
RTK-fix	793	677	29	41
RTK-float	12792	20424	3778	9504

エネルギー関数は、1 フレーム以下の精度での画像と GPS の同期を扱えないため、画像の取得と GPS 測位の時刻には、0.5 フレーム分のずれが生じる可能性がある。そこで、実際の円柱のサイズとしては、表 2(b) に示した水平方向、高度方向それぞれの値に、0.5 フレームの間に人が動く距離として 37mm と 5mm を足した値を用いる。(注1)

### 3.2 動画と GPS 測位値を用いたカメラ位置・姿勢推定

#### 3.2.1 実験条件

本実験では、ワイドレンズ (Sony VCL-HG0758) を取り付けたビデオカメラ (Sony DSR-PD-150) を用いて約 70m の距離を移動しながら、手持ちで撮影した動画 (解像度 720 × 480 画素、プログレッシブ撮影、15 fps、1110 フレーム) と RTK-GPS 受信機 (TOPCON GR-3) から得られた 1 秒間隔で計測された GPS 測位値を入力として用い、カメラ位置・姿勢の推定を行った。ただし、GPS 測位値は全て RTK-fix として取得され、これを本実験における真値として用いる。得られた GPS 測位値から後処理ソフトウェア (TOPCON Tools) を用いて、仮想的に GPS 衛星をマスクすることにより、精度の低い RTK-float のデータを作成し、これを手法の入力として用いた。

カメラと GPS 受信機は相対位置関係を固定し、カメラ座標系における GPS 受信機の位置  $d$  は、トータルステーションで計測した基準点の三次元位置を画像上で指定することによって得たカメラ位置・姿勢と、トータルステーションで計測した GPS 受信機の位置から算出した。カメラの内部パラメータはあらかじめ Tsai の手法 [18] によって求め、動画と GPS の測位値は手動で開始時刻を同期させた。また、式 (1) で用いられる GPS 測位位置に関するペナルティの重みを  $\omega = 10^{-8}$  とし、狭区間最適化を行うフレーム間隔を  $k = 15$ 、使用フレームを  $l = 500$  に設定した。

#### 3.2.2 比較対象

本実験では、提案手法の有効性を確認するため、以下の手法とカメラ位置の推定精度を比較する。

手法 A: GPS 測位値を用いず、動画のみから推定を行う手法。

手法 B: GPS の誤差として正規分布を仮定し、GPS の測位精度を考慮しない従来手法 [12]。

手法 C: GPS の誤差として正規分布を仮定し、正規分布の分散を GPS の測位精度を考慮して変化させる従来手法 [13]。

手法 D: 提案手法において、GPS の誤差範囲を回転楕円体とした場合。

ただし、ここでは手法 A によって推定されるカメラ位置・姿勢を GPS 測地座標系で評価するため、第 1 フレームから第 30 フレームまでのカメラ位置・姿勢を画像上で三次元位置が既知の基準点を指定することによって与えた。また、これらの基準点は手法 B, C, D, および提案手法において特徴点の追跡処理の初期化のためにも使用した。

手法 D は、2.1.2 で定義した GPS 測位位置に関するペナルティを決定するための、GPS 測位位置を中心とした一定の範囲を円柱から回転楕円体に変更したものであり、手法 D との比較によって、範囲の形が推定精度に与える影響を検証する。具体的には、一定の範囲を回転楕円体とした場合の GPS 測位位置に関するペナルティ項  $\Psi_i$  を以下のように定義する。

$$\Psi_i = \left( \frac{1}{\alpha(p)} \sqrt{x_i^2 + \left( \frac{\alpha(p)}{\beta(p)} y_i \right)^2 + z_i^2} \right)^{2n} \quad (10)$$

ただし、 $\alpha, \beta$  はそれぞれ水平方向、高度方向の回転楕円体の径の 1/2 である。円柱の場合と同様に、GPS 受信機の位置が測位位置を中心とした回転楕円体の中に存在する場合には  $\Psi_i$  が 0 に近い値となり、回転楕円体の外となる場合には非常に大きな値となる。また、回転楕円体のパラメータ  $\alpha(p), \beta(p)$  は、3.1.2 で決定した円柱のパラメータと同様に、GPS 測位値の定点観測の結果から決定する。ただし、回転楕円体の体積が最も小さくなるように、円柱の場合と同数の測位値を外れ値として除去する。決定した回転楕円体のパラメータを表 3 に示す。

#### 3.2.3 カメラ位置推定精度の定量的評価

提案手法によって推定されたカメラ位置・姿勢と特徴点の三次元位置を図 4 に示す。図中の曲線は推定されたカメラパスを、錐体は 60 フレーム毎のカメラの姿勢を表し、錐体中の画像はそのフレームにおける入力画像を表す。同図から、カメラパスは滑らかに推定されていることが分かる。また、目視によって環境の三次元構造が概ね正しく復元されていることを確認した。なお、本実験では PC (Xeon 2.66GHz, 8core) を用い、処理 (A) ~ (C) に約 60 分、処理 (D) に約 102 分を要した。

各手法における、水平方向および高度方向の GPS 受信機の位置の推定結果と GPS の測位位置をそれぞれ図 5, 6 に示す。また真値との比較による、フレームごとの位置誤差を図 7 に、位置誤差の平均、標準偏差、最大値を表 4 に示す。図 7 より、動画のみを用いる手法 A は

(注1): 後述する実験では、人が手に持ったカメラで歩行中に撮影された 15fps の動画を用いる。このとき、人の歩行速度を時速 4 km とすれば 0.5 フレーム間の移動量は 37mm となる。また、上下運動の振幅を 15 cm とすれば 0.5 フレーム間の移動量は 5mm となる。

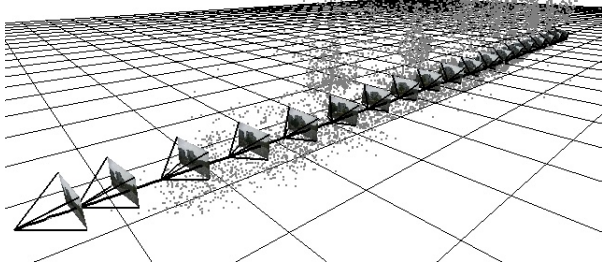
表 3 回転楕円体のパラメータ [mm]

測位精度	$\alpha$	$\beta$
RTK-fix	25	55
RTK-float	4125	11366

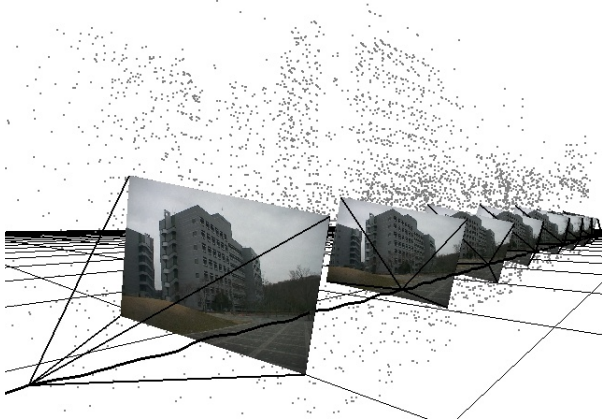


表 4 位置誤差の比較 [mm].

手法	平均	標準偏差	最大
手法 A	2109.6	915.8	3036.2
手法 B	1787.0	1497.3	4742.5
手法 C	894.7	582.2	1680.7
手法 D	405.5	286.7	901.6
提案手法	406.0	261.8	844.2



(a) 全体



(b) 一部拡大

図 4 推定されたカメラ位置・姿勢と特徴点の三次元位置

手法 C, D, および提案手法と比較して, 位置誤差の増加率が大きいことから, カメラ位置・姿勢の推定結果が蓄積誤差に大きく影響されていることが分かる。これは, 手法 A が動画のみを用いており, 蓄積誤差を抑える外部指標を持たないためである。また, 図 5, 6 より, 手法 B は測位精度の低い RTK-float の測位位置に大きく影響されていることが分かる。これは, 手法 B が GPS の測位精度を考慮せず, GPS の測位誤差は小さいと仮定しているためである。これに対して, 手法 C, D, および提案手法は GPS の測位精度を考慮することにより, 手法 B より RTK-float による誤差の影響が小さくなっている。これは, GPS の測位精度を考慮することで RTK-float の影響を抑えつつ, 動画の初期フレームと最終フレーム付近に存在する RTK-fix の測位位置を用いて蓄積誤差を最小化しているためである。

ただし, 手法 C では, 手法 D, および提案手法と比較して RTK-float の影響がより大きく, 推定されるカメラ位置が RTK-float の測位位置に引き寄せられている。これは, 手法 C が GPS の測位誤差として平均値 0 の正規分布を仮定しているのに対し, 短時間における RTK-float の平均測位位置が真値から外れているためであると考え

られる。手法 D, および提案手法では RTK-float の測位位置に大きく影響されることなく, カメラ位置が推定されている。

表 4 から分かるように, 本実験では, 手法 D, および提案手法は従来手法 C に対して 2 倍程度の精度向上が見られる。また, GPS の測位位置を中心とした一定の範囲を回転楕円体とした手法 D と, 円柱とした提案手法を比較して, カメラ位置の推定精度にほぼ差がないことから, 本実験においては範囲の形は推定精度に大きく影響しないことが分かる。

#### 4. ま と め

本報告では, 三次元形状復元や自由視点画像生成などオフラインで高精度なカメラ位置・姿勢を必要とするアプリケーションへの適用を前提とし, GPS と動画の併用によるカメラ位置・姿勢推定において, GPS 受信機の真の位置が GPS の測位位置を中心とする GPS の測位精度に依存した一定の大きさの円柱内に存在することを仮定し, この仮定に基づくエネルギー関数を最小化することで, GPS の測位精度が低い場合にも精度良くカメラ位置・姿勢の推定が可能な手法を提案した。実験では, 屋外環境で取得した動画と GPS の測位値を用いてカメラ位置・姿勢を推定し, 従来手法との定量的な推定精度の比較を行った。実験の結果, GPS の測位精度が低い場合において, 提案手法は従来手法より精度良くカメラ位置を推定できることを確認した。このことから, 本手法は, 屋外環境の三次元形状復元や自由視点画像生成などへ適用できると考えられる。

今後の課題として, GPS 測位に含まれる外れ値の自動除去手法の開発, 動画と GPS の同期の自動化, GPS の解の種類以外の測位精度に影響を与える要因の考慮などがあげられる。外れ値の除去と同期の自動化は, GPS から得られる情報と動画から得られる情報の整合性を, エネルギー関数の値の大きさを評価尺度として検証することで実現できると考えられる。また, 測位精度に影響を与える要因の考慮として, GPS の測位におけるマルチパスなどの環境に依存する誤差を, 動画からの推定によって得られる環境の三次元構造を用いることにより見積もることが考えられる。

謝辞 本研究の一部は科学研究費補助金(基盤研究(A), No.19200016), 総務省・戦略的情報通信研究開発推進制度(SCOPE)による。

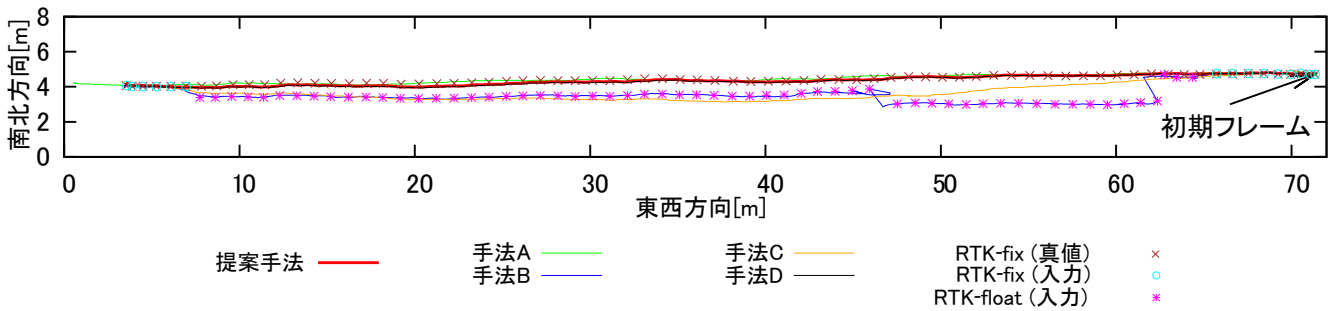


図 5 推定された GPS 受信機の位置 (水平方向)

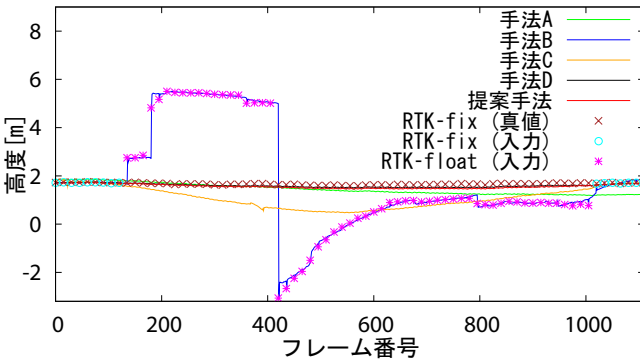


図 6 推定された GPS 受信機の位置 (高度方向)

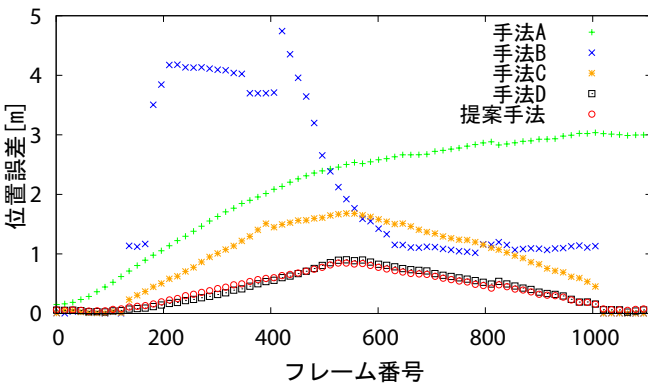


図 7 フレームごとの位置誤差

## 文 献

- [1] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. Kim, P. Merrell, et al.: "Detailed real-time urban 3D reconstruction from video," *Int. J. of Computer Vision*, Vol. 78, No. 2-3, pp. 143–167, 2008.
- [2] S. Knorr, M. Kunter and T. Sikora: "Super-resolution stereo- and multi-view synthesis from monocular video sequences," *Proc. Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 55–64, 2007.
- [3] A. Stafford, W. Piekarski and B. H. Thomas: "Implementation of god-like interaction techniques for supporting collaboration between outdoor AR and indoor tabletop users," *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 165–172, 2006.
- [4] 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.
- [5] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: "Bundle adjustment - a modern synthesis," *Proc. Int. Workshop on Vision Algorithms*, pp. 298–372, 1999.
- [6] G. Klein and D. Murray: "Parallel tracking and mapping for small AR workspaces," *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 225–234, 2007.
- [7] 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.
- [8] 佐藤, 神原, 横矢, 竹村: "マーカと自然特徴点の追跡による動画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. J86-D-II, No. 10, pp. 1431–1440, 2003.
- [9] G. Bleser, H. Wuest and D. Stricker: "Online camera pose estimation in partially known and dynamic scenes," *Proc. Int. Symp. on Mixed and Augmented Reality*, pp. 56–65, 2006.
- [10] M. Agrawal and K. Konolige: "Real-time localization in outdoor environments using stereo vision and inexpensive GPS," *Proc. Int. Conf. on Pattern Recognition*, Vol. 3, pp. 1063–1068, 2006.
- [11] D. Schleicher, L. M. Bergasa, M. Ocana, R. Barea and E. Lopez: "Real-time hierarchical GPS aided visual SLAM on urban environments," *Proc. Int. Conf. on Robotics and Automation*, pp. 4381–4386, 2009.
- [12] 横地, 池田, 佐藤, 横矢: "特徴点追跡と GPS 測位に基づくカメラ外部パラメータの推定", *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. 47, No. SIG5(CVIM13), pp. 69–79, 2006.
- [13] T. Anai, N. Fukaya, T. Sato, N. Yokoya and N. Kochi: "Exterior orientation method for video image sequences with considering RTK-GPS accuracy," *Proc. Int. Conf. on Optical 3-D Measurement Techniques*, Vol. I, pp. 231–240, 2009.
- [14] 安田: "GPS 技術の展望", *電子情報通信学会論文誌 (B)*, Vol. J84-B, No. 12, pp. 2082–2091, 2001.
- [15] 白井: "高精度衛星測位技術の動向と応用", *電子情報通信学会誌*, Vol. 92, No. 9, pp. 768–774, 2009.
- [16] C. Harris and M. Stephens: "A combined corner and edge detector," *Proc. Alvey Vision Conf.*, pp. 147–151, 1988.
- [17] 栗田, 坂上: "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*, pp. 65–70, 2000.
- [18] R. Y. Tsai: "An efficient and accurate camera calibration technique for 3D machine vision," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 364–374, 1986.