

位置の連続性を考慮したバンドル調整による 動画像と GPS からのカメラ位置・姿勢推定

糸 秀行[†] 武富 貴史[†] 佐藤 智和[†] 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: †{hideyuki-k,takafumi-t,tomoka-s,yokoya}@is.naist.jp

あらまし 本報告では、広域屋外環境における動画像と GPS を併用したカメラ位置・姿勢推定において、推定されたカメラ位置が不連続になるという問題を解決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、バンドル調整の誤差関数に GPS 測位に関する項を追加したエネルギー関数を最小化することにより、蓄積誤差を抑えたカメラ位置・姿勢の推定を可能としている。しかし、GPS と動画像の計測周期の違いにより、GPS 測位を利用できるフレームと利用できないフレームが存在し、これらのフレーム間で推定されるカメラ位置が不連続になるという問題がある。これに対して本研究では、従来提案されているエネルギー関数に、新たに位置の連続性に関する項を追加することで、推定されたカメラ位置が不連続になるという問題を解決する。

キーワード カメラ位置・姿勢推定, Structure from Motion, GPS, バンドル調整

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での利用が可能である。近年これらの分野では、広域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おり、様々な手法が提案されている [1, 2]。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 [3] や Visual SLAM 法 [4] 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。

このような誤差蓄積の問題に対し、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, 5-10] 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 [5, 6] は、三次元位置が既知の基準点 [5] 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 [6] 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, 7-10] 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。このような特徴から、本研究では、GPS と動画像を併用したカメラ位置推定手法に着目する。

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, 7, 8] とバンドル調整 (bundle adjustment) を用いる手法 [9, 10] に大別できる。カルマンフィルタを用いる手法 [1, 7, 8] は、動画像と GPS の情報を高速に融合できるため、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。しかし、過去の状態と現在の観測から逐次的に現在の状態を求めるというフィルタの設計から、入力全体での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 [9, 10] 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エネルギーを追加することで、動画像と GPS の情報を融合する。横地ら [9] は、GPS の測位位置に関するエネルギーとして、GPS の測位位置と GPS 受信機の推定位置の間の距離を用いている。この手法では、定義したエネルギー関数を最小化することで、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 測位の信頼度は 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、GPS の測位位置と GPS 受信機の推定位置の間の距離をエネルギーとして用いるだけでは、特に GPS 測位の信頼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 [10] は GPS 受信機から得られる信頼度情報に応じて GPS の測位位置に関するエネルギーの重みを変化させることにより、測位の信頼度を考慮する手法を提案している。しかし、これらバンドル調整を用いる GPS と動画像情報の融合手法には、GPS と動画像の計測周期の違いにより、GPS 測位に直接対応するフレームと対応しないフレームが存在し、これらのフレーム間で図 1 に示すように、推定されるカメラ位置に不連

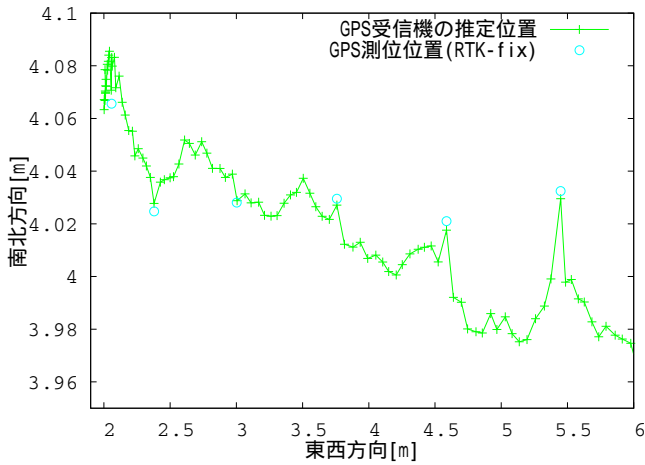


図1 Anai らの手法 [10] によって推定された GPS 受信機の位置の例 (縦軸と横軸でスケールが異なることに注意)

続が生じるという問題がある。また、最小化されるエネルギー関数には局所解が多数存在し、初期値によっては、間欠的に得られる信頼度の高い GPS 測位の情報を推定結果全体に反映することが難しい場合が存在する。

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想定し、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 [10] を基礎とし、(1) 推定されるカメラ位置が不連続になるという問題、(2) 初期値によって信頼度の高い GPS 測位の情報を推定結果全体に反映することが難しいという問題、を解決する。具体的には、従来提案されているエネルギー関数に、新たにカメラ位置の連続性に関する項を追加したエネルギー関数を定義し、このエネルギー関数を最小化することでカメラ位置・姿勢を推定する。また、エネルギー関数を最小化する前に、推定結果全体を信頼度の高い GPS 測位位置へ当てはめることで良い初期値を生成し、これを用いて最小化を行う。

以下、2 節では、カメラ位置の連続性を考慮したエネルギー関数の定義および最適化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. 位置の連続性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。

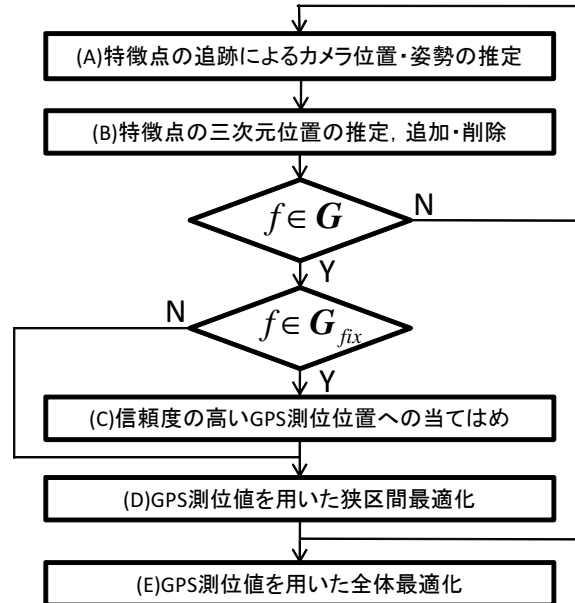


図2 提案手法の処理の流れ

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 [10] を基礎とし、手法 [10] で提案された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に加え、本研究で新たに定義するカメラ位置の連続性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 2 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 [5]。ここで、現フレーム (第 f フレーム) において、GPS 測位が得られた場合 (GPS 測位時刻に対応する動画像のフレームの集合 G に f が含まれる場合)、GPS の測位値を用いた狭区間最適化 (D) 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。ただし、信頼度の高い GPS 測位が得られた場合 (信頼度の高い GPS 測位に対応するフレームの集合 G_{fix} に f が含まれる場合) には、GPS 測位位置に推定結果全体を当てはめることで、良い初期値を生成し、その後最適化を行う。これらの処理 (A) ~ (D) を初期フレームから最終フレームまで繰り返すことで、逐次的に蓄積誤差を抑えたカメラの位置・姿勢を得た後、最後に動画像全体での最適化処理を行う (E)。以下では、まず、本研究で新たに提案するカメラ位置の連続性を考慮したエネルギー関数について述べる。次に、処理 (C) ~ (E) について詳述する。

2.1 カメラ位置の連続性を考慮したエネルギー関数の定義

本研究では、従来のバンドル調整による動画像と GPS の融合手法 [10] で用いられている特徴点の再投影誤差と GPS 測位位置に関するエネルギーにカメラ位置の連続性に関するエネルギーを追加したエネルギー関数を最小

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 フレームにおける特徴点 j の再投影誤差 Φ_{ij} と i フレームにおける GPS 測位位置に関するエネルギー Ψ_i 、カメラ位置の連続性に関するエネルギー Θ_i を用いて以下のように定義する．

$$E = \frac{1}{\sum_{i \in F} |P_i|} \sum_{i \in F} \sum_{j \in P_i} \Phi_{ij} + \frac{\omega}{|G|} \sum_{i \in G} \Psi_i + \frac{\lambda}{|F|} \sum_{i \in F} \Theta_i \quad (1)$$

ただし、 F は動画のフレームの集合、 G は GPS 測位時刻に対応する動画のフレームの集合、 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を表す．また、 ω は GPS 測位位置に関するエネルギーの重み係数、 λ はカメラ位置の連続性に関するエネルギーの重み係数である．以下、再投影誤差 Φ_{ij} 、GPS 測位位置に関するエネルギー Ψ_i 、カメラ位置の連続性に関するエネルギー Θ_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 Φ_{ij} を以下のように定義する．

$$\Phi_{ij} = \mu_j (q_{ij} - \hat{q}_{ij})^2 \quad (2)$$

ただし、 q_{ij} は、第 i フレームにおける特徴点 j の画像上での検出座標、 \hat{q}_{ij} は特徴点 j の三次元位置を画像上へ投影した座標である．また、 μ_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 [5]．

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

ただし、 σ_j^2 は特徴点 j の再投影誤差の分散を表す．

2.1.2 GPS 測位位置に関するエネルギー

従来手法 [10] は、GPS の測位位置に関する誤差として、GPS 測位位置と GPS 受信機の推定位置の間の距離を用い、以下のような GPS 測位位置に関するエネルギーを定義している．

$$\Psi_i = \frac{1}{\tau_i^2} (g_i - \hat{g}_i)^2 \quad (4)$$

$$\begin{pmatrix} \hat{g}_i \\ 1 \end{pmatrix} = \begin{pmatrix} R_i & t_i \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d \\ 1 \end{pmatrix} \quad (5)$$

ただし、 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、 \hat{g}_i は第 i フレームの GPS 測地座標系における GPS 受信機の推定位置を表す． R_i 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

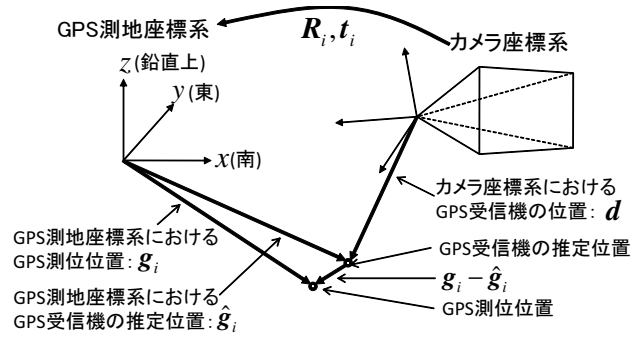


図3 GPS 測位位置に関するエネルギー

回転、 t_i は GPS 測地座標系におけるカメラ位置、 d はカメラ座標系における GPS 受信機の位置を表す．

また、 τ_i は GPS 測位の信頼度に応じた重みであり、以下のように定義される．

$$\tau_i = \begin{cases} 1 & ; i \in G_{fix} \\ \tau_{float} & ; otherwise \end{cases} \quad (6)$$

ただし、 G_{fix} は信頼度の高い GPS 測位に対応するフレームの集合である．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 [11, 12]．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 (RTK-fix、RTK-float) 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、本研究では、信頼度の高い GPS 測位として RTK-fix 解の GPS 測位を用いる．また、信頼度の低い GPS 測位の重み τ_{float} は GPS 測位の定点観測によって得られる測位位置の RMS から実験的に決定する．

2.1.3 カメラ位置の連続性に関するエネルギー

本研究では、推定されるカメラ位置の連続性に関するエネルギーを GPS 測地座標系におけるカメラ位置 t_i を用いて以下のように定義する．

$$\Theta_i = (t_i - t_{i+1})^2 \quad (7)$$

2.2 GPS 測位値を用いた最適化

2.2.1 最適化の範囲

GPS 測位値を用いた最適化には、図 2 に示したように、逐次処理において範囲を限定して行う狭区間最適化 (D) と、逐次処理の終了後、動画全体を対象として行う全体最適化 (E) がある．

処理 (A)、(B) は、動画上の情報のみを考慮するため、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生じる．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、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 (E) 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 (D) では、逐次処理において、第 $(f-l)$ フレームから現フレーム (第 f フレーム) を用いて、式 (1) で定義したエネルギー関数 E を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 (E) では、処理 (A) ~ (D) 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エネルギー関数 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。

2.2.2 エネルギー最小化手法

提案手法では、エネルギー関数 E の最小化に Levenberg-Marquardt (LM) 法を用いる。ただし、LM 法において必要となる E のヤコビ行列が疎になることから、疎な行列を考慮した LM 法 (sparseLM) [13] を用いて最小化を行う。

2.2.3 信頼度の高い GPS 測位位置への当てはめ

エネルギー関数 E はパラメータが非常に多く、局所解が多数存在すると考えられる。そこで、エネルギー関数 E を最小化する前に、推定されているカメラ位置・姿勢全体を回転、並進、スケール変換の 7 パラメータによって信頼度の高い GPS 測位位置へ当てはめることで、良い初期値を得た上で最小化を行う。具体的には、以下の式によって求められる回転 R_{fit} 、並進 t_{fit} 、スケール変換 s_{fit} によって、現時点までに推定されているカメラ位置・姿勢と特徴点の三次元位置を変換する。

$$\begin{aligned} & (R_{fit}, t_{fit}, s_{fit}) = \\ & \operatorname{argmin}_{R, t, s} \sum_{i \in G_{fix}} \left(\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_i & st_i \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d \\ 1 \end{pmatrix} - \begin{pmatrix} g_i \\ 1 \end{pmatrix} \right)^2 \end{aligned} \quad (8)$$

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、従来手法との推定精度の定量的な比較を行うことで、提案手法の有効性を示す。

3.1 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 の距離を東から西へ直進しながら、手持ちで撮影した動画像 (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、GPS 受信機の推定位置 \hat{g}_i と比較することで位置推定精度の定量的な評価を行う。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、図 4, 5 に示すような、信頼度の低い RTK-float のデータを含む GPS 測位値を作成し、これをすべての比較手法の入力として用いた。表 1 に、利用した RTK-GPS 受信機のカタログ仕様を示す。また、特徴点の追跡処理の初期化のため、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。

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 およびカメラの内部パラメータはあらかじめキャリブレーション済みで既知とした。また、動画像と GPS の測位値は手動で開始時刻を同期させ、式 (1) で用いられる GPS 測位位置に関するエネルギーの重みを $\omega = 10^{-8}$ 、位置の連続性に関するエネルギーの重みを $\lambda = 10^{-8}$ 、狭区間最適化の使用フレームを $l = 1110$ に設定した。式 (6) で用いられる信頼度の低い GPS 測位の重み τ_{float} は GPS の定点観測結果より $\tau_{float} = 107.4$ とした。

3.2 比較対象

本実験では、提案手法におけるカメラ位置の連続性に関するエネルギーおよび信頼度の高い GPS 測位位置への当てはめの有効性を確認するため、以下の手法と位置推定精度を比較する。

手法 A: 従来手法 [10]

手法 B: 提案手法においてカメラ位置の連続性に関するエネルギーのみを用い、GPS 測位位置への当てはめを行わない場合

手法 C: 提案手法において GPS 測位位置への当てはめのみを行い、カメラ位置の連続性に関するエネルギーを用いない場合

手法 D: GPS 測位位置への当てはめを行い、カメラ位置の連続性に関するエネルギーを用いる場合 (提案手法)

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

