動画像と GPS の併用によるカメラ位置・姿勢推定における GPS 測位精度の考慮

Consideration of GPS positioning accuracy for camera parameter estimation from GPS and video images

粂 秀行 武富 貴史 佐藤 智和 横矢 直和

Hideyuki Kume, Takafumi Taketomi, Tomokazu Sato and Naokazu Yokoya 奈良先端科学技術大学院大学 情報科学研究科

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

ABSTRACT This paper proposes a method for estimating extrinsic camera parameters without accumulative errors using video images and position data acquired by GPS. Conventional methods assume that GPS error is normally distributed. However, GPS error generally includes bias errors, and the accuracy of the estimated position becomes lower as this bias becomes larger. In order to resolve this problem, instead of the assumption for the normal distribution error, the proposed method assumes that true GPS position exists in the certain range whose size depends on the GPS positioning accuracy. Concretely, the proposed method estimates camera parameters by minimizing the error function that summing up the reprojection error and the penalty term for GPS positioning.

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.このよ うな目的に対して,従来,撮影された動画像のみからカメ ラの位置・姿勢を推定する手法 [1,2]が提案されている. Kleinら [2]は特徴点の再投影誤差の最小化処理 (bundle adjustment)と特徴点の追跡処理を並列に実行することに より,実時間での高精度なカメラ位置・姿勢推定を実現し ている.しかし,動画像のみを用いる手法では推定結果に 誤差が蓄積するという問題や,スケール情報が得られない 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[3,4,5,6,7]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[3,4]では, 三次元位置が既知の基準点[3]やシーン中の一部の領域に 対するワイヤーフレームモデル[4]を利用する.これらの 手法では基準点やモデルの三次元位置情報を用いることで 蓄積誤差とスケールの不定性の問題を解決している.しか し,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.GPSを併用する手法[5,6,7]では,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, 事前準備なしに, GPS 測地座標系における蓄積誤差のない カメラの位置・姿勢を得ることができる.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 [5,6] と bundle adjustment を用いる手法 [7] に大 別できる.

カルマンフィルタを用いる手法 [5,6] は低い計算コスト で最小二乗規範に基づく入力データの統合を行うことが できる.しかし,カルマンフィルタは直前の状態と現在の 入力情報から現時刻の状態を逐次的に予測するため,既 に決定された過去の状態を修正することは難しい.一方, bundle adjustment を用いる手法 [7] では,一般的な動画像 に対する bundle adjustment 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 報を融合している.この手法では,一定の範囲内の全ての パラメータに対する最適化処理を行うことで,現在の入力 情報を用いて過去の推定結果を修正することができる.こ れにより,カルマンフィルタに対して,より精度の高い推 定を可能とする.しかし,カルマンフィルタを用いる手法 を含め,これら GPS を併用する従来手法では,GPS の測 位誤差を平均値が0の正規分布と仮定しており,特に GPS の測位精度が低い float 解や differential 解が得られた場合 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 から大きく外れる誤差が生じた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.

他方,歩行動作に基づくデッドレコニングやオドメトリ とGPSを併用することで,人やロボットの位置・姿勢を推 定する手法[8,9]も,誤差が蓄積する点および,相対的な移 動しか得られないという点で上記の研究に関連があるが, これらの手法はカルマンフィルタを用いて融合を行ってい るため,動画像への適用を考えた場合にはやはり上述のカ ルマンフィルタを用いる手法[5,6]と同様の問題がある.

本研究では,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・姿勢推定において低い人的コス トで高精度な絶対位置情報を推定可能な bundle adjustment による GPS と動画像情報の融合手法を基礎とし,手法の 改良を行う.具体的には,GPS の測位誤差を正規分布と仮 定せず,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ことを仮定し,この仮 定に基づく誤差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 し,GPS の測位精度が低い場合にカメラ位置・姿勢の推定 精度が低下することを防ぐ.

以下,2節では,GPSの測位精度を考慮した誤差関数の 定義とカメラ位置・姿勢推定手法について詳述し,3節で は,屋外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・姿勢推定結果を示す.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, 文献 [7] と同様にカメラと 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.提案手法は逐次処理と全体最適化処理からなり, 両処 理に GPS の測位値を用いることで誤差の蓄積を抑える.

提案手法の処理の流れを Fig.1 に示す.まず,フレーム 毎の処理として,(A)特徴点の追跡によるカメラ位置・姿 勢推定,(B)特徴点の三次元位置の推定,追加・削除を行 う.次に処理(C)では,一定フレーム間隔 k 毎に本研究で



Fig.1 Flow diagram of the proposed method

新たに定義する GPS の測位位置と特徴点の再投影誤差に 基づく誤差関数を最小化することで,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.拡張現実 感などの逐次的なカメラの位置・姿勢を必要とするアプリ ケーションではこの処理により得られる結果を用いる.オ フラインでの処理が可能な三次元形状復元や自由視点画像 生成を目的とする場合には,逐次処理の完了後に動画像全 体で誤差関数を最小化する(処理 D).以下では,まず処理 (C),(D) で用いる誤差関数を定義し,その最適化手法につ いて述べる.次に処理(A)~(D) について詳述する.

2.1 誤差関数の定義

本研究では,最適化処理に用いる誤差関数を,第*i*フレームにおける特徴点位置に関する誤差 Φ_i と GPS 測位位置に 関する誤差 Ψ_iを用いて以下のように定義する.

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

ただし、 ω は重み係数、F は動画像のフレームの集合、 F_g は GPS 測位時刻に対応する動画像のフレームの集合を表 す.誤差関数 E を最小化することで、動画像全体でカメラ の位置・姿勢と特徴点の三次元位置を最適化する.以下、 特徴点位置に関する誤差 Φ_i と GPS 測位位置に関する誤差 Ψ_i について詳述する.

2.1.1 特徴点位置に関する誤差

文献 [7] と同様に,動画像からのカメラ位置・姿勢推定 において一般的に用いられる再投影誤差を用いて,特徴点



Fig.2 Error associated with GPS position

の位置に関する誤差 Φ_i を以下のように定義する.

$$\Phi_i = \frac{\mu_i}{|\mathbf{S}_i|} \sum_{j \in \mathbf{S}_i} \kappa_j (\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij})^2$$
(2)

ただし, S_i は第 i フレームの画像上で検出される特徴点の 集合, μ_i は各フレームに対する重み係数であり GPS 測位 が得られたフレームに大きな値を設定する. κ_j は逐次処理 で得られた特徴点 j の信頼度を表す.また, x_{ij} , \hat{x}_{ij} はそ れぞれ, 第 i フレームにおける特徴点 j の画像上での検出 座標と,特徴点 j の三次元位置を画像上へ投影した座標を 表す.

2.1.2 GPS 測位位置に関する誤差

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,Fig.2 に示すように,真 の GPS 受信機の位置は 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 とし,GPS 測位位置に関する誤差 Ψ_i を以下のように定義 する.

$$\Psi_{i} = \left(\frac{1}{r(p)}\sqrt{x_{i}^{2} + y_{i}^{2}}\right)^{2n} + \left(\frac{1}{h(p)}z_{i}\right)^{2n}$$
(3)
$$\begin{pmatrix} x_{i} \\ y_{i} \\ z_{i} \\ 1 \end{pmatrix} = M_{i}d - g_{i}$$
(4)

ただし, M はカメラ座標系から GPS 測地座標系への変換 行列, d はカメラ座標系における GPS 受信機の位置, g は GPS 測地座標系における GPS の測位位置, 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 Ψ_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 Ψ_iを微分可能な形で定義することで,最適化処理における 勾配計算を可能としている.

2.2 GPS の測位精度を考慮したカメラ位置・姿勢推定

本節では, Fig.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(A)

ここでは,入力画像上で特徴点を追跡することによりカ メラの位置・姿勢を推定する.ここで得られたカメラの位 置・姿勢は後述する特徴点の三次元位置の推定(B)やGPS 測位値を用いた最適化(C)における初期値として用いら れる.

具体的には,まず Harris オペレータ[10]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.次に前フレーム 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.仮の追跡位置から,LMedS 基準を用い たロバスト推定[11]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.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づけを行 い,追跡位置を確定する.最後に追跡された特徴点の画像 上の位置と三次元位置から再投影誤差を最小化することに よりカメラの位置・姿勢を得る.

2.2.2 特徴点の三次元位置の推定,追加・削除(B)

ここでは,特徴点の三次元位置の推定と追加・削除を行う.まず,各フレームにおいて推定されたカメラの位置・ 姿勢と特徴点の画像上の位置から,再投影誤差を最小化す ることで特徴点の三次元位置を推定し,毎フレーム更新す る.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自動で特徴点の追加・削除 を行なう.

2.2.3 GPS 測位値を用いた最適化 (C),(D)

処理 (A),(B) によって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.そこで,処理 (C),(D) では式 (1) で定義した誤差関数 *E* を,勾配法によっ

Table.1 Maximum error of each positioning accuracy [mm]

測位精度	(a) 全データ		(b) 外れ値除外後	
	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
単独測位	254085	356620	19872	29665

て最小化することで, 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(C)で は k フレームごとに,現フレームから l フレーム前までの フレームを用いて誤差関数 E の最小化を行うことにより, 狭区間での最適化を行う.これにより,逐次的な推定にお いて誤差の蓄積を抑える.また,処理(D)では全てのフ レームについて逐次的な処理が終了した後,全てのフレー ムを用いて誤差関数 E を最小化する.

3 実験

本節では,まず GPS を固定して観測を行い,式(3)で 用いる測位精度ごとの位置計測の誤差範囲を決定する. 次に,カメラと GPS を移動させながら取得した動画像と GPS の測位値を用いて従来手法[7]との比較を行い,提案 手法の有効性を確認する.

3.1 測位精度ごとの位置計測の誤差範囲の決定

RTK-GPS 受信機 (TOPCON GR-3) を固定して約5時間 の連続測位を行った.Table.1(a)は,得られた全てのRTKfix データの平均位置を真値とし,測位精度ごとに真値か らの誤差の最大値を算出したものである.Table.1(a)より, 測位精度ごとに誤差の最大値が大きく異なることが確認で きる.同表(b)には測位精度ごとに,水平方向,高度方向 それぞれについて誤差が大きい5%の測位値を除外した結 果を示す.外れ値を除外することにより,誤差の最大値が 大幅に小さくなっていることから,GPSの測位値には誤差 が極端に大きい外れ値が少数含まれていることが確認でき る.以降の実験では,GPS 測位の外れ値が何らかの手法に より除外できるものとし,Table.1(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.

3.2 動画像と GPS 測位値を用いたカメラ位置・姿勢推定 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 240m

の距離を移動しながら撮影した動画像 (解像度 720 × 480 画素,プログレッシブ撮影,15fps,3106 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てカメラ位置・姿勢 の推定を行い,動画像のみを用いる手法,GPS の誤差とし て正規分布を仮定する従来手法 [7] との比較を行った.た だし,動画像と GPS の測位値は手動で開始時刻を同期さ せ,カメラと GPS 受信機の相対位置関係は手動で計測し た値を用いた.カメラの内部パラメータはあらかじめ Tsai の手法 [12] によってキャリプレーションし,その値を用 いた.また,処理の初期化のため第 50 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点を指定す ることによって与え,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 = 150に設定した.

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 Fig.3, Fig.4 に示す.た だし,提案手法では2248 フレーム,従来手法では1835 フレーム,動画像のみを用いた場合では2231フレームで 特徴点の追跡に失敗し,以降のカメラ位置・姿勢は推定さ れていない.これらの図より,動画像のみを用いて推定を 行った場合では蓄積誤差の影響によりカメラパスが大幅に 縮んでいることがわかる.また,従来手法ではこのような, 誤差が蓄積するという傾向は見られないが, GPS の測位精 度が低い場合 (RTK-float,単独測位) において,カメラ位 置が GPS 測位位置に引きよせられていることが確認でき る.これに対して,提案手法では,GPSの測位精度が低い 場合においても,このような問題が生じていないことが分 かる.また,提案手法において,全体最適化を行うことに より,逐次的な推定結果に対して,より滑らかなカメラパ スが推定されていることが確認できる.

提案手法によって推定されたカメラ位置・姿勢を用いて 仮想物体を動画像に合成したマッチムーブの結果を Fig.5 に示す.ワイヤーフレームはトータルステーションを用い て計測したビルの角の三次元座標をもとに描画されてい る.同図 (a),(b) は,それぞれ逐次処理,全体最適化によ り得られた結果を用いたものである.ワイヤーフレームが おおむね正しい位置に描画されていることから推定された カメラ位置・姿勢に大きな計測誤差が含まれていないこと が確認できる.しかし,逐次的な推定結果において 686 フ レームと 691 フレームの間でワイヤーフレームが大きく動



Fig.3 Result of estimated GPS positions(horizontal direction)

いていることが分かる.これは 691 フレームにおいて測位 精度の高い GPS 測位 (RTK-fix) が得られ, 蓄積誤差が急激 に小さくなったためだと考えられる.これに対して,全体 最適化を行った結果では,ワイヤーフレームが実環境中の 一定の位置に描画され,カメラ位置・姿勢が滑らかに推定 されていることが確認できる.

4 まとめ

本稿では,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,この仮定に基づく誤差関数 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手法を提案し た.実験では,提案手法を用いることで,GPS の測位精度 が低い場合においても,GPS の測位位置に大きく依存する ことなくカメラ位置が推定できることを確認した.今後の 課題として,GPS 測位に含まれる外れ値の除外手法の開 発,GPS 測位値を用いた初期化などが挙げられる. 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.

参考文献

- M. Pollefeys, R. Koch, M. Vergauwen, B. Deknuydt and L. Van Gool: "Three-dimensional scene reconstruction from images," Proc. SPIE, Vol. 3958, pp. 215–226, 2000.
- [2] 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.
- [3] 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. J86-D-II, No. 10, pp. 1431–1440, 2003.
- [4] 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.
- [5] 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.
- [6] D. Schleicher, L. M. Bergasa, M. Ocana, R. Barea and



Fig.4 Result of estimated GPS positions(altitude)



681 フレーム

(a) 逐次処理

691 フレーム

686 フレーム

696 フレーム





Fig.5 Result of match move by the proposed method

E. Lopez: "Real-time hierarchical GPS aided visual SLAM on urban environments," Int. Conf. on Robotics and Automation, pp. 4381–4386, 2009.

- [7] 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.
- [8] M. Kourogi, N. Sakata, T. Okuma and T. Kurata: "Indoor/outdoor pedestrian navigation with an embedded GPS/RFID/self-contained sensor system," Proc. Int. Conf. on Artificial Reality and Telexistence, pp. 1310–1321, 2006.
- [9] K. Ohno, T. Tsubouchi, B. Shigematsu, S. Maeyama and S. Yuta: "Outdoor navigation of a mobile robot between build-

ings based on DGPS and odometry data fusion," Proc. Int. Conf. on Robotics and Automation, pp. 1978–1984, 2003.

- [10] C. Harris and M. Stephens: "A combined corner and edge detector," Proc. Alvey Vision Conf., pp. 147–151, 1988.
- [11] 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.
- [12] 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.