

因子分解法とバンドル調整による全方位画像列からの 形状および位置姿勢の同時推定手法

松久 亮太[†] 川崎 洋^{††} 小野晋太郎[†] 阪野 貴彦[†] 池内 克史[†]

[†] 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{††} 埼玉大学 埼玉県さいたま市桜区下大久保 255

E-mail: [†]{matsuhisa,onoshin,vanno,ki}@cvl.iis.u-tokyo.ac.jp, ^{††}kawasaki@cgv.ics.saitama-u.ac.jp

あらまし シーンの形状およびセンサの位置姿勢推定は、VR やロボティクスなど様々な分野で必要とされており、これまで多くの研究が行われてきた。例えば屋外などでシーンの形状および位置姿勢を知る方法として、レンジセンサに GPS やジャイロセンサなどを併用するシステムが提案され、実際に利用されている。しかし、このような複数のセンサによる手法では微小な移動や傾きなどを正確に推定することが出来ず、精密なシーンの形状復元や位置姿勢の推定などは容易ではない。一方、センサとしてカメラを用いて、撮影した画像列から形状およびカメラの位置姿勢の同時推定を行う Structure from Motion に関する研究が試みられている。この方法ではピクセル精度による高精度な推定が可能ではあるものの、初期値依存性が高く、またノイズなどの影響を受けやすいため、推定結果が不安定になるといった問題がある。本研究では、Structure from Motion 手法に関して、全方位画像列を入力として、初期値を因子分解法で得て、これをバンドル調整法で精度向上させることで、形状および位置姿勢を同時に精度良く頑健に推定する手法を提案する。

キーワード Structure from Motion, 全方位画像列, 因子分解法, バンドル調整法

Structure from Motion for Omni-directional Images by using Factorization and Bundle Adjustment Method

Ryota MATSUHISA[†], Hiroshi KAWASAKI^{††}, Shintaro ONO[†], Atsuhiko BANNO[†], and Katsushi
IKEUCHI[†]

[†] Institute of Industrial Science, the University of Tokyo
4-6-1 Komaba Meguro-Ku, Tokyo, 153-8505 JAPAN

^{††} Saitama University 255 Shimo-Okubi, Sakura-ku, Saitama City, Saitama 338-8570, JAPAN

E-mail: [†]{matsuhisa,onoshin,vanno,ki}@cvl.iis.u-tokyo.ac.jp, ^{††}kawasaki@cgv.ics.saitama-u.ac.jp

Abstract Recently, the estimation of a 3D shape of a scene as well as the position and the orientation of the sensor is extensively researched, especially for VR and Robotics systems. To achieve this, the system consists of laser range sensor, GPS and Gyro sensor has been proposed, actually constructed and used. However, it is usually difficult to estimate a precise and detailed estimation of those parameters because of limitation of ability of each sensors. On the other hand, Structure from Motion(SFM) method is widely known for such purposes and the method can estimate those parameters accurately as pixel resolution. However, SFM method is usually unstable dependent on initial parameters and noise. In this paper, we propose a SFM method for omnidirectional image sequences using both factorization and bundle adjustment method to achieve high accuracy and robustness.

Key words Structure from Motion, Omnidirectional Images, Factorization, Bundle Adjustment

1. はじめに

シーンの形状およびセンサの位置姿勢推定は、バーチャルリ

アリティ(VR)における実シーンと仮想シーンの合成 [1] や、ロボティクスにおける Simultaneous localization and mapping (SLAM) [2] など様々な場面において必要とされており、これま

で多くの研究が行われてきた。例えば屋外などにおいて、シーンの形状および位置姿勢を知る方法として、レンジセンサをロボットに搭載し、そのロボットの位置姿勢を GPS やジャイロセンサなどを用いて獲得し、それらデータを融合し利用するシステムがこれまで提案され、実際に利用されている。しかし、一般に用いられているレーザレンジセンサの精度は広域を対象としたものだと通常数 cm 程度の誤差がある。また、GPS やジャイロセンサも非常に高精度なものも存在するものの、例えば実際に撮影した画像と、CAD などの 3 次元データを、1 ピクセル単位で位置あわせするほどの精度は実現されていない。このため、これらセンサを単純に併用する手法では、精密なシーンの形状復元や位置姿勢の推定などを実現することは困難である。

一方で、センサとしてカメラを用いて、撮影した画像列から形状およびカメラの位置姿勢の同時推定を行う Structure from Motion に関する研究が試みられている。この方法ではピクセル精度による推定が可能のため、非常に高精度な推定が実現できる可能性があるものの、最適化するパラメータが多く評価関数も非線形であるため、初期値依存性が高く、またノイズなどの影響を受けやすく、推定結果が不安定になりがちである。このため、実用として使用するためには多くの問題があり現在も多くの研究が行われている。最近では、3 次元位置の既知なマーカや GPS など外部センサを併用することで、精度の良い初期値を得て、形状と位置姿勢の高精度な同時推定を行う研究例 [3] [4] も見られるが、手動でのマーカの設置やセンサ群の高精度な同期が必要であるなど簡易に構築できるシステムとはなっていない。もし、カメラで得られた画像のみから、形状と位置姿勢を精度良く推定することが出来れば、システムをより簡易に構築出来、利用価値も高い。また、外部センサと併用する場合においても、外部センサがうまく動作しない場合などにおいて、高精度な補間に利用でき、システムの実際の運用に大きく寄与することができる。

そこで、本研究では、全方位画像列を入力とし、初期値を因子分解法で得て、これをバンドル調整法で精度向上させることで、形状および位置姿勢を同時に精度良く頑健に推定する Structure from Motion 手法を提案する。また、入力として全方位画像列を用いることで、全方位のうち最も位置姿勢に敏感な特徴量が常に最適化に寄与するため、推定精度が向上しこれにより解の安定性向上が期待される。

2. 関連研究

Structure from Motion において、入力に全方位画像を用いることで解の安定性の向上が期待できるため、このような研究が近年多く行われている。以下では、それら手法の代表例について簡単に紹介する。

全方位画像の広範囲な視野を活かし、単一カメラの移動前後の画像を利用したステレオ計測による推定 [5] が提案されている。しかし、単一カメラにおけるステレオ計測では、カメラの移動距離がそのまま基線長となるため、カメラ間の距離が近接している場合には精度良く推定できない。そのため、本論文のように細かい間隔でカメラの位置姿勢を高精度に推定すること

は難しい。また、Structure from Motion の一手法である因子分解法を魚眼全方位画像へ適用した手法 [6] では、光軸まわりの回転と 2 次元移動のみ前提とし、位置姿勢推定を行っている。自由度がこの 3 自由度に限定される限りにおいては極めて有効であると考えられるが、本論文のように自由度が比較的大きい場合には、初期値としての利用が考えられるものの、そのまま利用することは難しい。また、6 自由度全ての位置姿勢を推定する手法が提案されている [7]。ここでは、初期推定には自然特徴点を利用するが、最終的な推定の際に 3 次元位置が既知な基準マーカを利用しており、提案手法が既知のマーカを一切用いない点で前提が異なる。ただし、このようなマーカを用いることが出来る場合においては現実的な手法である。

3. 提案手法の概要

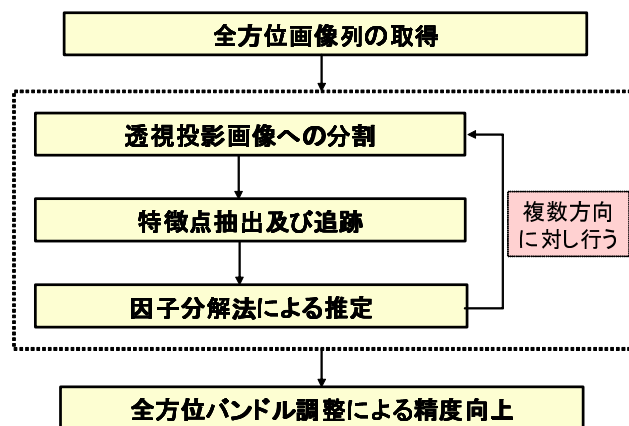


図 1 形状および位置姿勢推定手法の概要

本手法のアルゴリズムの概要を図 1 に示す。まず全方位カメラシステムを利用し取得した全方位画像列の各フレームから一部分を分割し、一般的なカメラ画像に近い透視投影画像に変換する。得られた画像列から特徴点抽出および追跡を行い、因子分解法によって 3 次元座標とカメラ位置姿勢を推定する。ここで推定されるパラメータは全方位画像を分割・変換して得られた透視投影画像を撮影したと考える仮想的なカメラである。以上の処理を分割部分を変えながら複数回を行い、複数方向のパラメータを推定する。因子分解法により得られた 3 次元点とカメラ位置姿勢は、カメラモデルの相違などにより通常誤差を含むため、これらを初期パラメータとし、バンドル調整法による最適化により精度向上を図る。その際、複数方向の全てのパラメータを同時に用いることで、よりロバストな推定を行うことができる。以下に実際の処理について述べる。

3.1 全方位画像列の取得

全方位画像列の取得には Point Grey Research 社 [8] の全方位カメラ Ladybug2 を搭載し (図 2)、走行しながら撮影を行った。このカメラでは内部に 6 つの CCD が近接して設置されており、全周の約 75% の視野をカバーする合成全方位画像を容易に得ることができる。

全方位画像を得る方法としては他にも 1 台のカメラと曲面鏡



図2 全方位カメラシステム Ladybug2 とその車載例

を使用する方法が、また因子分解法に入力する画像としては

- (1) 全方位画像そのもの
- (2) 全方位画像が合成される前の各 CCD の画像 (Ladybug2 の場合)

(3) 全方位画像を任意に分割して透視投影にした画像等が考えられるが、本実験では Ladybug2 の合成画像に 3 を適用することとした。これは、画像の解像度が高く、特徴点の追跡時間が長く取れ、また既存の因子分解法がそのまま使えるといった利点を考慮したものである。

3.2 画像分割方法

本手法では画像上の特徴点座標を入力として形状およびカメラ位置姿勢推定を行う。しかし Ladybug2 によって得られた全方位画像は座標軸がそれぞれ天頂角と方位角を表す球面座標系となっている。そのため、この画像の特徴点座標をそのまま通常の因子分解法の入力として用いることは出来ない。そこで、全方位画像の一部の範囲に注目し、一般的なカメラである透視投影モデルの画像へ変換を行い入力とする。



図3 Ladybug2 による全方位画像を透視投影画像に変換した例

得られた変換画像列 (図3) に対し特徴点追跡を行うが、各フレームで全方位画像上の同じ範囲を分割変換しても、特徴点を長いフレームに渡って追跡し続けることは出来ない。また、短いフレームで追跡を行えたとしても、画像上の見えに大きな差異が無い場合、得られる奥行き情報も少なく、正確な推定を行うことが出来ない。そこで本論文では、長く特徴点を追跡できるように、フレーム毎に分割範囲をずらしながら透視投影変換画像列を作成する手法を提案する。これにより、ちょうど通常

のカメラを回転させながら撮影した場合と同様の画像を得る事が出来 (図4), 因子分解法の解の安定に寄与する。これは全方位画像を用いる大きなメリットの一つであり、提案手法の利点でもある。回転方法としては、全方位画像の時点で特徴点追跡を行い、可能な限り多くの特徴点を含むよう自動的に行うのが理想であるが、今回は直進運動を仮定し、手動により回転のパラメータを与えることで同じ景観を写し続けるようにした。

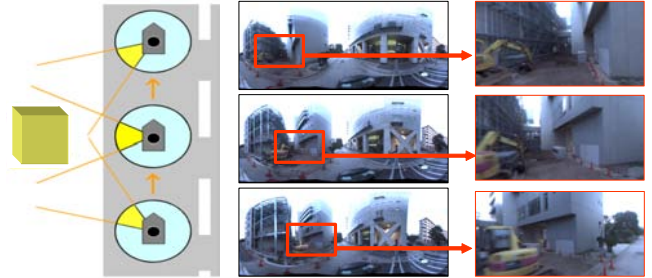


図4 全方位画像の分割範囲の変更

3.3 特徴点抽出および追跡

3.2 章で得られた透視投影画像列の特徴点抽出および追跡を行う。まず初期フレームの Harris オペレータ [9] を特徴点抽出を行う。得られた特徴点のオプティカルフローを Lucas-Kanade 法 [10] によって計算し、追跡を行う。ただしフローが見つからなかった場合、その特徴点の追跡はその時点で終了する。本稿で初期推定に利用する因子分解法では、全てのフレームに軌跡が残っている特徴点のみを利用できないため、最終フレームまで追跡を行えた特徴点をラベル付けて確保しておく。さらに今回は、こうして残った特徴点のうち、誤りがないものを手動でさらに選別を行った。ビデオレートで撮影した画像列での特徴点のトラッキングは高精度で行えるため、通常このような手動選択は不要であるが、今回の因子分解法の実装では僅かなノイズにより結果に大きな誤差が発生することがあったためこのような選別を行った。大きな誤差が発生理由の解明に関しては今後の課題である。

3.4 因子分解法による位置姿勢推定

因子分解法 [11] は、多視点画像から 3 次元形状とカメラ位置姿勢を推定する手法である。追跡した特徴点座標の軌跡に対し特異値分解を行う事で、形状とカメラ運動を同時に求めることができる。特に今回は透視投影下を想定した因子分解法 [12] を用いた。

因子分解法によって推定される解は、3 次元形状の重心を原点とする重心座標系となっている。そのため、前節で推定した複数方向の因子分解法の解を、すべて共通の座標系へ変換する処理を行う。変換は推定したカメラ位置姿勢を基準として利用する。変換後の座標系は、初期フレームのカメラ位置を原点とし、天頂方向を Y 軸、カメラの進行方向を Z 軸とするワールド座標系とした。

3.5 バンドル調整法による精度向上

全方位のデータを有効利用するために左右それぞれの方向で仮想透視変換カメラを設定し、それぞれに因子分解法を適用す

ることが考えられる。また、因子分解法は利用する特徴点が全てのフレームに渡って画像上に存在する必要があるため、一度に長い区間を対象とせず、短い区間に区切ってそれぞれに適用すると安定した処理が実現できる。ところが、これらの分割処理により、一台の全方位カメラによって撮影された特徴点の形状およびカメラの位置姿勢推定が、各方向と各区間毎に独立した状態で復元される。そこで各方向による復元結果を一つにまとめ、全体の誤差が最小化するように非線形最適化（バンドル調整）を行う。推定パラメータは特徴点の3次元座標とそれぞれの全方位カメラの位置姿勢である。また最小化すべき誤差としては、入力として使用した特徴点の2次元座標と、上記パラメータ群を用いて透視投影を行った特徴点の2次元座標とのずれの総和を用いる（図5）。

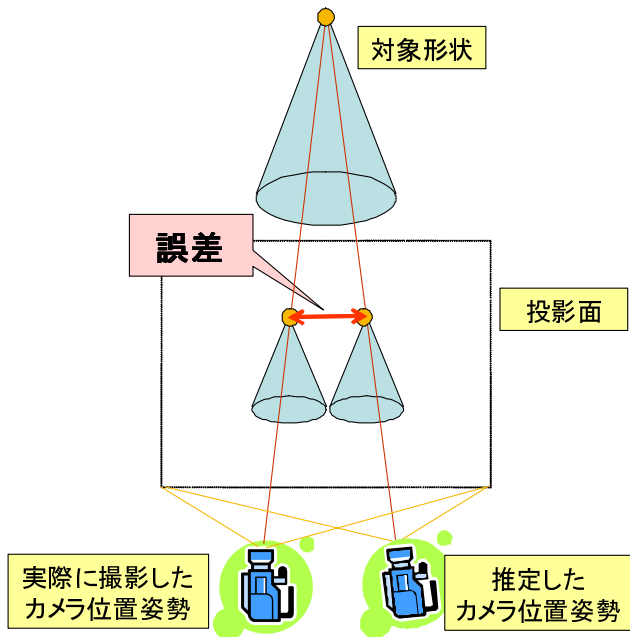


図5 バンドル調整にて最小化すべき誤差

最小化する目的関数は式(1)のようになる。 m と n はそれぞれ動画の全フレーム数と特徴点の数、 X は入力画像上の特徴点座標、 X' は因子分解法で推定された特徴点の3次元座標、 P は同じく推定されたカメラ座標系への変換行列である。ただし、 P は因子分解法の推定結果である重心座標系による位置姿勢パラメータを、ワールド座標系に変換した後のものである。

$$\min_{\hat{P}_i, \hat{X}'_j} \sum_{i=1}^m \sum_{j=1}^n (X_j - P_i X'_j)^2 \quad (1)$$

4. 実験

4.1 全方位シミュレーションデータによる実験

図6のような3次元点（黒点）とカメラ（赤点：位置，四角錐：姿勢）のシミュレーションデータを作成し、実験を行った。カメラは等速直線運動を仮定し、それぞれのカメラ位置から複数方向に対して透視投影（図7）を行い、それぞれの方向ごとに因子分解法で初期値の推定、およびその後の全方位でのバンドル調整法を適用した。この実験では、進行方向を軸に左右2

方向で別々に初期推定を行った。総フレーム数は10枚とし、透視投影による仮想的なカメラの姿勢は天頂角周りのみ変化させ、より多くの特徴点を全てのフレームに渡って追跡できるように回転を与えた。追跡した特徴点の数は左右共に44個であった。

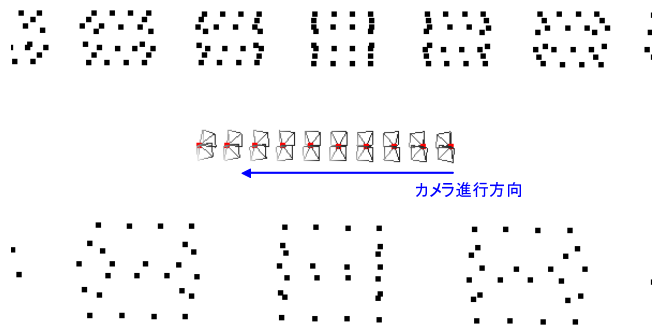


図6 全方位シミュレーションデータ

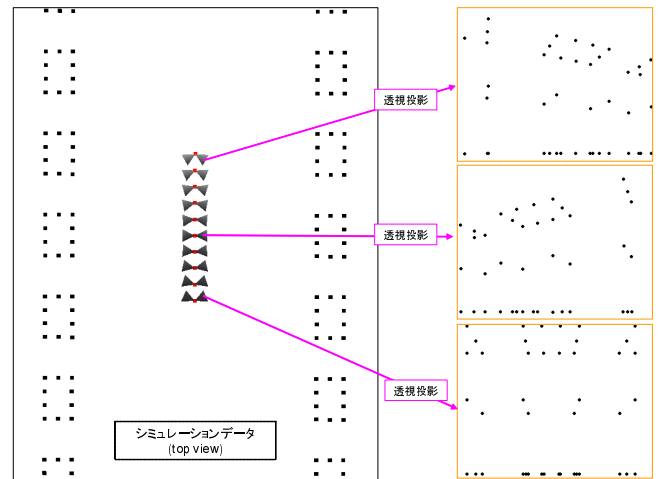


図7 全方位シミュレーションデータによる透視投影図の例

正解データ (Top View: 平行投影)	因子分解法のみ	因子分解法+バンドル調整

図8 全方位シミュレーションデータによる実験結果

図8は推定された3次元点とカメラ位置を上視点から見た結果である。因子分解法の時点では、左方向の形状に歪みが見られ、本来直進しているカメラの軌跡も曲がって推定されていることが分かる。これはカメラの軌跡が直線運動のみのため、このように奥行き情報が不足した場合において、単純な因子分解法ではカメラモデルがアフィンカメラではないことなどから、大きな誤差につながったものと考えられる。また、評価のためにカメラに回転を加えずに特徴点をトラッキングする手法での復元も試みたものの、十分な特徴点が得られなかったため解が

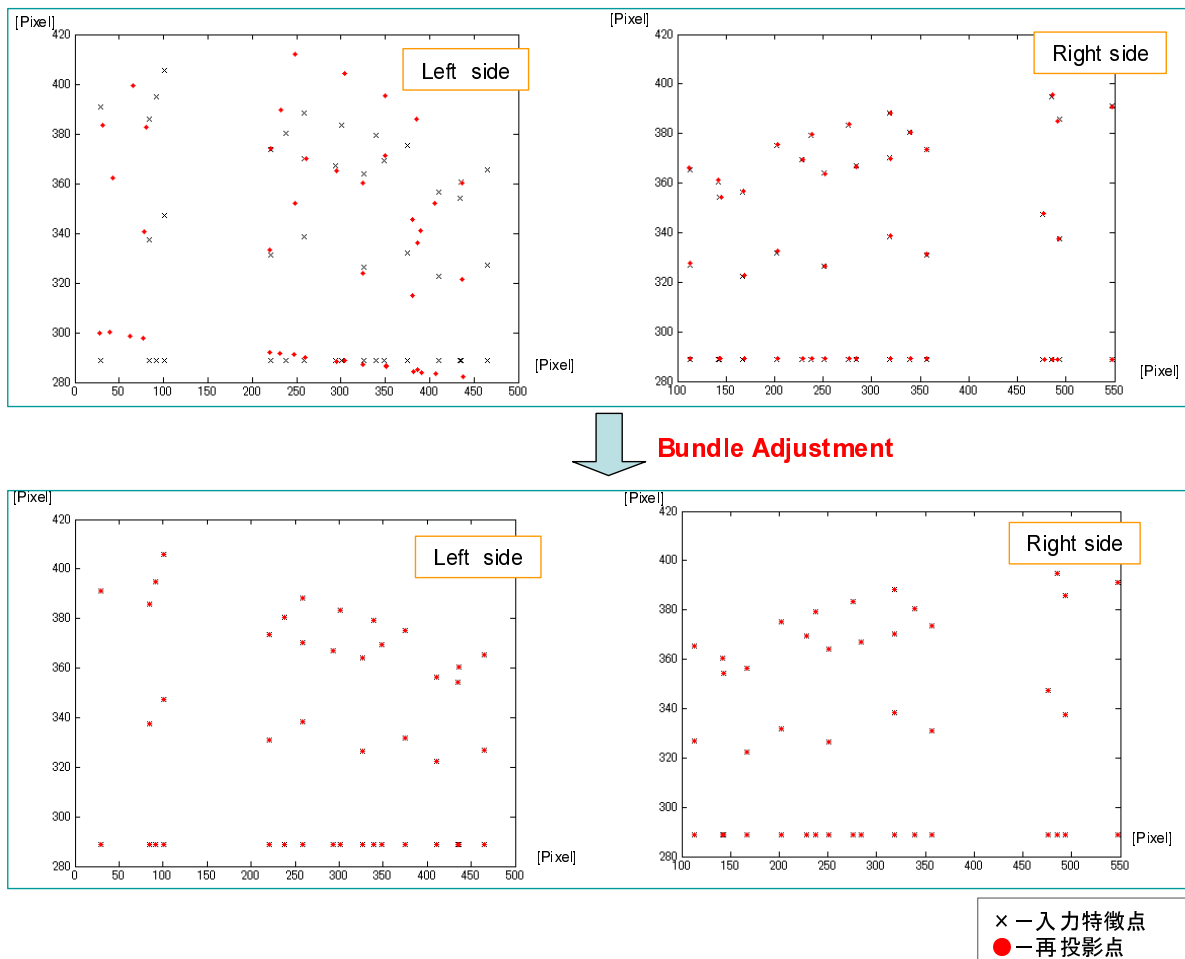


図 9 推定値の再投影点と入力特徴点の比較

縮退しプロット可能な解を得ることができなかった。このことより、提案したカメラに回転を加えて特徴点をトラッキングする手法により、歪みが大きく軽減され、初期として十分利用可能な解を得ることが出来、手法の有効性が確認できた。

また、バンドル調整後は形状の歪みもなくなり、ほぼ真値に等しい推定を行う事ができた。また画像に投影した座標の比較(図 9)でも、左右共に誤差が小さくなっており、複数方向同時最適化が有効であることが分かる。

4.2 屋外全方位実画像による実験

Ladybug2 によって取得した全方位画像列に対して実験を行った。図 10 は、画像を分割し、透視投影画像へ変換したイメージである。分割する範囲は、出来るだけ多くの特徴点がトラッキング可能なように以下のように定めた。まず、左方向は初期フレームが天頂角周りに左に 50 度回転した姿勢から、毎フレーム 9 度ずつ回転しながら、特徴点をトラッキングするようにし、次に右方向は初期フレームが天頂角周りに右に 40 度回転した姿勢から、毎フレーム 10 度ずつ回転することとした。このような設定で抽出した特徴点から、さらに手動で追跡する特徴点を左右共に 33 個に絞り込んだ。

図 11 は推定された 3 次元点とカメラ位置を上視点から見た結果である。因子分解法の時点では、両方向の形状に大きな歪みが見られた。しかし、バンドル調整後は形状の歪みはなくな



図 10 全方位画像列の透視投影画像への変換例

り、真値に近い推定を行う事ができた。図 12 は、少し分かりにくいですが、右上の図に示された特徴点が視点位置の変化に応じて、3次元空間中に推定されたカメラの位置姿勢と同時に示されたものである。またシミュレーション実験と同様、再投影による比較(図 13)でも、左右共に誤差が小さくなっており、複数方向同時最適化が有効であることが分かる。

今回、因子分解法による初期解がかなり歪んでいる原因としては、カメラの軌跡がほぼ直線運動であることも考えられるが、入力として利用した Ladybug2 の全方位画像にも一因があると思われる。Ladybug2 は計 6 台の CCD カメラによる画像を全方位画像として統合するシステムとなっているが、各々の CCD カメラの光学中心の一致は実現されておらず、画像に大きな歪みが含まれていることが知られている。これら 6 台の CCD カメラのキャリブレーションは既に幾つかの手法が提案されており、本手法における今後の課題である。



(a) 左側

(b) 右側

	バンドル調整なし	バンドル調整あり
Top view		
Back view		

図 11 全方位実画像による実験結果

4.3 連続区間による屋外全方位実画像による実験

また、先のフレーム区間より前のフレーム区間に対しても同様に推定を行った。その際、因子分解法まではそれぞれ別々の推定を行ったが、最終的な形状およびカメラの位置姿勢の推定は、全てのフレームを用いてバンドル調整を行った。図 14 上の緑色の特徴点は、先に推定を行ったフレーム間でも利用した共通の特徴点である。その結果、図 15 の結果が得られた。本来因子分解法ではスケーリングなどの自由度が残るものの、それぞれの形状の位置関係やスケールについても統一した座標系において推定することができたことが分かる。動画像では短い区分では精度良くトラッキングが行えることから、因子分解法も頑健な処理が期待できる。そうして得られた複数の初期推定結果を得て、これら全てをバンドル調整法で同時に高精度化することで、非常に長い距離についても精度の高い推定が可能で

ある。

5. まとめ

本論文では、全方位画像列を用いた精度の高い形状およびカメラ位置姿勢を同時推定する手法について述べた。

提案手法は、全方位画像を分割して因子分解法を行い、その解をまとめて最適化することで精度向上を図る。実際にシミュレーションデータと全方位カメラシステム Ladybug2 を用いて取得した全方位実画像列を用いて実験を行った。その結果、シミュレーションデータ実験ではほぼ真値と同じ推定結果を得ることが出来た。実画像での実験では、入力として使用した全方位画像の歪みや量子化誤差などにより若干のずれは生じたものの、ある程度真値に近い形状およびカメラ位置姿勢を推定することができたと考えられる。また、連続区間において別々に因子

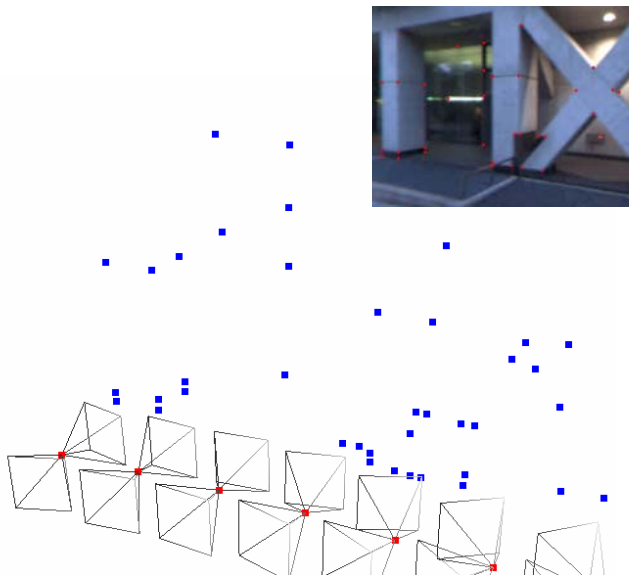


図 12 全方位実画像による実験結果 (右上・透視投影画像)

分解法を適用し、それらに含まれる共通な特徴点やカメラパスを用いて、バンドル調整による統合を行った。その結果、連続区間においても精度良く統一座標系における復元を実現することが出来た。今後は、全方位画像の歪みの除去を行うと同時に、さらに長い区間において復元を行う予定である。

謝辞

本研究は、東京大学生産技術研究所先進モビリティ連携研究センター (ITS センター)、朝日航洋 (株)、アジア航測 (株)、(株) デンソー、(株) トヨタマップマスターによる「次世代 ITS 計測車両の共同開発」プロジェクトの一環として行われました。

文 献

- [1] T. Kakuta, T. Oishi and K. Ikeuchi: “Shading and shadowing of architecture in mixed reality”, 4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 05) (2005).
- [2] H. Choset and K. Nagatani: “Topological simultaneous localization and mapping(slam): Toward exact localization without explicit localization”, IEEE Trans. on Robotics and Automation, pp. 125–137 (2001).
- [3] A. Ronald, B. Hoff, H. N. III and R. Sarfaty: “A motion-stabilized outdoor augmented reality system”, Proc. IEEE Virtual Reality, IEEE CS Press, pp. 252–259 (1999).
- [4] 大江, 佐藤, 横矢: “幾何学的位置合わせのための自然特徴点ランドマークデータベースを用いたカメラ位置・姿勢推定”, 日本バーチャルリアリティ学会論文誌, pp. 285–294 (2005).
- [5] 原田, 山下, 金子: “Structure from motion を用いた全方位カメラ搭載移動ロボットによる環境センシング”, 第 11 回ロボティクスシンポジウム講演予稿集, pp. 387–392 (2006).
- [6] 宮川, 若林, 荒川: “魚眼投影モデルに基づく画像系列からの平面運動と 3 次元形状の復元”, 電子情報通信学会論文誌, Vol.J87-D-II, No.5, pp. 1120–1132 (2005).
- [7] 佐藤, 池田, 横矢: “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, Vol.J88-D-II, No.2, pp. 347–357 (2005).

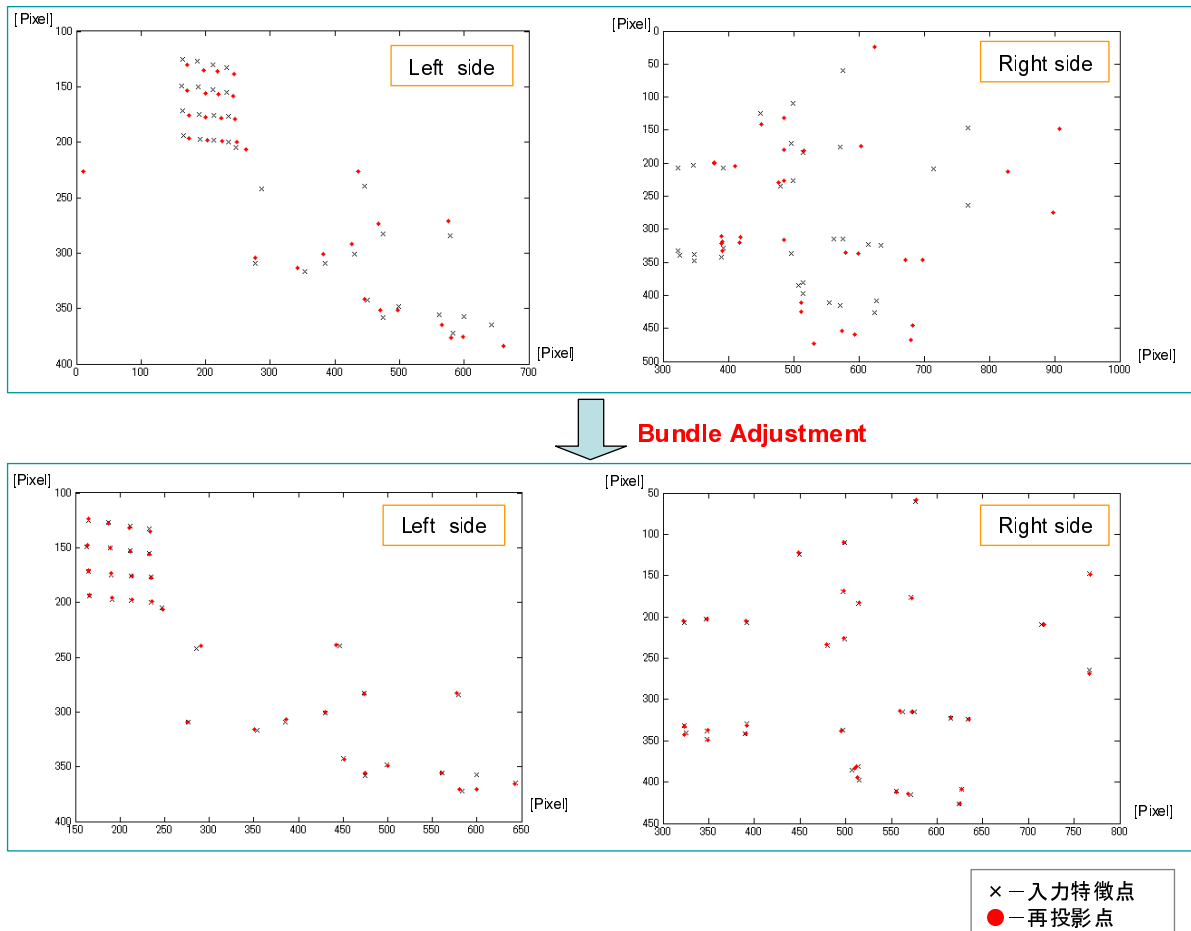


図 13 推定値の再投影点と入力特徴点の比較

- [8] Point Grey Research Inc. : Ladybug2
<http://www.ptgrey.com/products/ladybug2/>.
- [9] C. Harris and M. Stephens: "A combined corner and edge detector", Proc. Alvey Vision Conf., pp. 147-151 (1988).
- [10] B. Lucas and T. Kanade: "An iterative image registration technique with an application to stereo vision", Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), pp. 674-679 (1981).
- [11] C. Tomasi and T. Kanade: "Shape and motion from image stream under orthography: A factorization method", IJCV, **9**, pp. 137-189 (1992).
- [12] 阪野, 長谷川, 池内: "大型有形文化財計測のための気球搭載型レンジセンサ", CVIM, pp. 213-220 (2005).



(a) 左側

(b) 右側

図 14 バンドル調整適用前と後の再投影図

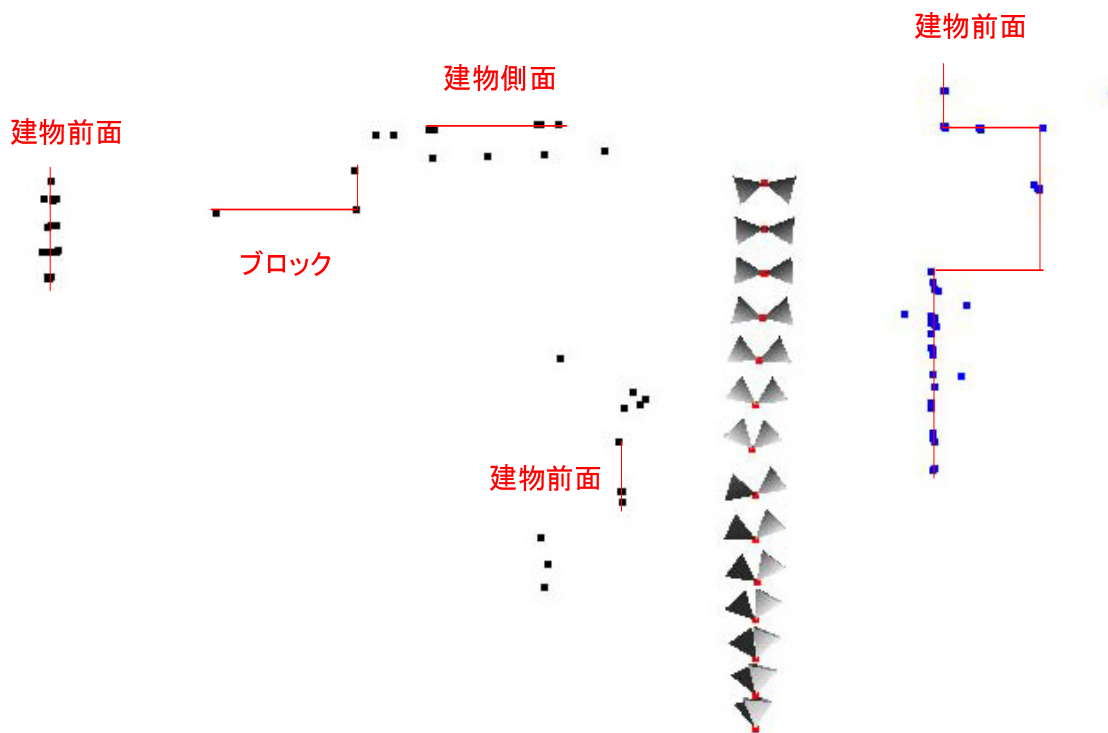


図 15 バンドル調整による推定データ統合