

ロバストバンドル調整を用いた 車載ステレオカメラのセルフキャリブレーション

深澤 至貴^{1,a)} 岡谷 貴之^{1,b)} 出口 光一郎^{1,c)} 木村 宣隆^{2,d)}

概要：車載ステレオカメラのセルフキャリブレーション，つまり車両が一般道を通常走行する際に得られる画像のみを用いたキャリブレーションの方法を述べる．具体的には，各カメラの内部パラメータおよびレンズ歪がオフラインのキャリブレーションで既知であるとし，ステレオカメラを構成する左右カメラの間の相対姿勢を推定する．この場合，自然のシーンを撮影した左右カメラの画像間でいかに外れ値のない点対応を得られるか，あるいは混入した外れ値を除去できるかが鍵となる．走行中の連続画像間および左右カメラの画像間で特徴点を追跡・対応付けてその軌跡を得，それらをすべて使用して完全な3次元復元(SfM)を行い，ロバストバンドル調整によって全体最適化をすることで，得た点対応に混入する外れ値を効果的に取り除き，高精度に目的パラメータを推定する．実験結果は，この方法がステレオカメラのキャリブレーションに有効であることを示した．

1. はじめに

近年，自動車の運転を支援し交通事故を未然に防ぐためのセンサとして，車載ステレオカメラが良く用いられるようになってきている．ステレオ視によりシーンの奥行き情報を得ることで，例えば前走車等を認識して衝突を未然に防いだり，道路の白線を認識して車が走行レーンから出てしまうことを防ぐなどの目的で利用されている．

そのためには，ステレオカメラのキャリブレーション—2台のカメラの画像と空間の幾何学的関係を与える各パラメータを知ること—が必要であり，しかもこれを高精度で行うことが求められる．知る必要のあるパラメータは，次の3種類に分けることができる．

P_1 : 2台のカメラそれぞれの内部パラメータ

P_2 : 両カメラの相対的な姿勢

P_3 : ステレオカメラの車両に対する姿勢

P_1 は，各カメラが持つ光学レンズの特性およびレンズの歪を指す．この P_1 に加えて P_2 が分かれば，左右カメラの画像間でマッチングを行うことでシーンのデプスを計算でき，さらに P_3 を知ることで上述のような応用が実現可能

となる．

基本的には P_1, P_2, P_3 いずれのパラメータも，そのカメラを搭載した車両を工場から出荷する前の段階で，何らかのキャリブレーション作業を行って決定することになる．それには，専用の器具，例えばキャリブレーション用のチャートやパタンを使うのが一般的である．これらの画像を車載ステレオカメラで実際に撮影し，その画像を用いて上の各パラメータを推定する．

本稿では，そのような専用の器具を使わないステレオカメラのセルフキャリブレーションの方法を示す．セルフキャリブレーションとは，車両が一般道路上を普通に走行している状態にあるとき，カメラが撮影する自然シーンの画像を元に上述のパラメータをオンラインで推定することである．

このようなセルフキャリブレーションには2つの意義がある．一つは，上述の工場出荷前のキャリブレーションの作業あるいは工程が不要になることである．もう一つは，たとえ出荷前キャリブレーションを行ったとしても，走行時の振動やその他の経年変化により，各パラメータが変化する可能性があり，その際の再キャリブレーションが容易に行えることである．ステレオカメラシステムを，車両の工場出荷後に据え付けることも可能になる．

本稿では， P_1, P_2, P_3 のパラメータのうち， P_2 のみを対象に，これをセルフキャリブレーションで推定する問題を考える． P_1 は，専用のチャートやパタンを用いないセルフキャリブレーションでは，最も推定が難しいパラメータで

¹ 東北大学
Tohoku University

² (株)日立製作所
Hitachi, Ltd.

a) fukasawa@fractal.is.tohoku.ac.jp

b) okatani@fractal.is.tohoku.ac.jp

c) kodeg@fractal.is.tohoku.ac.jp

d) nobutaka.kimura.fd@hitachi.com

ある。レンズ歪はともかく、焦点距離や特に画像中心を正確に求めるのは一般に難しいことが知られている。また、車載状態でのカメラの運動軌跡は直線上や平面上に拘束された状態に近く、臨界条件 (critical configuration) による縮退 (全パラメータを一意に決められない) の問題もある ([10], [11])。幸い P_1 は、左右の各カメラごとに別々に (取り外した状態でも) 実行でき、それには Zhang の方法 [12] など手軽な方法がいくつかある。

P_2 と P_3 を比べると、それぞれの推定は独立に行える上、 P_2 の推定がより難しく、重要である。 P_2 が得られていれば P_3 は、例えば車両が直進していることを画像外の情報源から得ることで、比較的容易に推定できる。なお P_2 の自由度は厳密には 6 自由度だが、ここではカメラ間の距離は既知とし、残りの 5 自由度のみを定めることを考える。

今、 P_1 が既知であるとしているので、 P_2 の 5 自由度の決定はカメラ間の基本行列を定めることとほぼ等価であり、左右カメラの画像間で最低 5 点以上の対応点があれば計算できる。したがって、今考えている自然シーンを対象としたセルフキャリブレーションでは、対応点をいかに高精度に得られるかが最大の課題となる。

自然シーンの画像から特徴点を得る場合、まず第一にその位置決め精度の問題がある。つまり、キャリブレーション用のパターンを用いる場合のそれに対して、その精度は一般に悪くならざるを得ない。ただしその位置決め誤差にバイアスがなければ (つまり期待値がその空間の点の画像上の位置に一致するならば)、大量の画像および特徴点を使うことでこの不足を補うことができるだろう。今考えている問題設定では、車両の走行中好きなだけデータを得られるから、オフラインキャリブレーションでは行い得ない大量のデータを使える。

自然シーンの画像を用いる場合のもう一つの問題は、画像間の対応付けに混入する外れ値であり、こちらがより深刻である。 P_2 を決めるための左右の画像間の対応点を得る最も単純な方法は、走行中の画像列から左右の画像ペアを複数抜き出し、それぞれ独立に画像間対応を求める方法であるが、この場合外れ値の除去は左右の画像間のエッジポラ条件のみで行うことになり、その外れ値除去性能は最小である。より良い方法は、連続画像における特徴点の軌跡を求め、時間軸方向の情報も利用することである。

時間軸の情報を利用する方法は何通りか考えられる。例えば左右それぞれのカメラの連続画像上で軌跡を求め、軌跡どうしを左右のカメラ間で対応付ければ、2, 3, 4 視点の多重線形拘束 (multi-linear constraint) [6] を用いることができ、これを RANSAC 等のロバスト推定の枠組みに取り入れることで、より強力に外れ値を除去できる。ただしこの方法は、(収束が保証されない非線形計算に頼らず外れ値除去を行える点で魅力的だが) 利用できる画像系列データのごく一部を使っているに過ぎず、全体で見てデー

タの情報量を最大限に引き出した推定を行っているとは言えない。例えば、いつも正しくインライアと外れ値を区別できるとは限らない。

これに対し、本稿ではまとまった数の画像系列 (約数百視点) を用い、得られる特徴点軌跡を使って 3 次元復元 (SfM) を行い、そこで全体最適化を行う方法を考えた。ここではロバストバンドル調整を使って、全体最適化の枠組みの中で外れ値除去を行う。このやり方では、外れ値除去のために、上述の多重線形拘束を包含する最も一般化された強力な拘束が自動的に適用される。多重線形拘束を使う場合と違って、最適化の計算が収束するか否かが問題となるが、外れ値の識別に関わる最適化パラメータを段階的に切り替えながら最適化を行う方法を述べる。

2. 関連研究

車載ステレオカメラのセルフキャリブレーションはいくつかの研究がある [4], [13], [14], [15]。中でも Dang らの研究 [4] が最も関連が深い。彼らは、車両走行時の画像系列上で追跡したの特徴点軌跡を対象に、拡張カルマンフィルタ (EKF) を使って、カメラパラメータをオンラインで推定する方法を述べている。彼らの方法は、Visual SLAM を単眼カメラで行う MonoSLAM [5] や類似研究 [1], [3] と同じ定式化に基づく。EKF を用いるのは、時間軸に沿って観測が増えたとき、一定の近似を行うことで解くべき問題のサイズを一定に抑えつつ、推定精度を確保するためである。

しかしながら、SfM/SLAM に EKF を用いると、観測モデルの非線形性によって推定精度が低下することが知られている。EKF は観測モデルを都度、線形近似し、また状態変数の事後分布をガウス分布で近似する方法であるから、推定精度を維持するには、これらの近似精度が担保されることが必要条件である。この条件が満たされない場合、特に観測モデルの非線形性が無視できないとき、EKF はしばしばいわゆる一貫性 (consistency) を失い、具体的には状態変数の不確かさ (共分散) を実際よりいつも小さく見積もってしまうことが知られている [7]。文献 [2] では、EKF-SLAM で実際に一貫性が失われることと、その原因の考察が示されている。

もう一つの方法はバンドル調整を用いて完全最適化を行うことである。SfM/SLAM へのアプローチとして、このフィルタリングとバンドル調整の精度および計算時間を比較した研究が [9] にあるが、後者が望ましいと結論されている。本研究はこの議論にも依拠している。

3. 問題の定式化

3.1 ステレオカメラの幾何モデル

左右のカメラは互いに剛体結合されているものとする。左右のカメラ間の相対姿勢は、左カメラ座標系から右カメラ座標系の次の座標変換によって表現できる。

$$\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t} \quad (1)$$

ただしカメラ座標系とは、両カメラともに投影中心を原点とし z 軸を光軸とする座標系であるとする。この相対姿勢を表すパラメータを 4×4 行列 \mathbf{T} によって

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (2)$$

と表す。

車両が走行する間、左右のカメラから画像が次々に得られる。画像を得た瞬間（左右のカメラは同期がとれているとする）のステレオカメラの視点を $i = 1, \dots$ によって表す。 $i = 1$ の左カメラのカメラ座標系を基準座標系にとり、そこから i 番目の視点における左カメラのカメラ座標系への座標変換を

$$\mathbf{x}_i = \mathbf{R}_i \mathbf{X} + \mathbf{t}_i \quad (3)$$

と書き、姿勢変化のパラメータを 4×4 行列

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (4)$$

で表す。なお $\mathbf{T}_1 = \mathbf{I}_4$ （単位行列）である。基準座標系から i 番目の視点における右カメラへの座標変換は

$$\mathbf{T}'_i = \mathbf{T}\mathbf{T}_i \quad (5)$$

で表せる。左辺の行列から回転と並進成分を取り出して、基準座標系から右カメラのカメラ座標系への座標変換は

$$\mathbf{x}'_i = \mathbf{R}'_i \mathbf{X} + \mathbf{t}'_i \quad (6)$$

のようになる。

左右のカメラのレンズ歪は次のような標準的なモデルで表す。左カメラのカメラ座標系での空間の点を $\mathbf{x} = [x, y, z]$ と書くとき、 $u \equiv x/z$ 、 $v \equiv y/z$ として、歪を受けた座標を

$$\bar{u} = u(1 + k_1 r^2 + k_2 r^4) + 2p_1 uv + p_2(r^2 + 2u^2) \quad (7a)$$

$$\bar{v} = v(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2v^2) + 2p_2 uv \quad (7b)$$

とする。ただし $r^2 = u^2 + v^2$ である。歪パラメータは $[k_1, k_2, p_1, p_2]$ である。右カメラも同様で、その歪パラメータを $[k'_1, k'_2, p'_1, p'_2]$ とする。最終的な画像上の投影位置は、左右のカメラの内部パラメータを \mathbf{K}, \mathbf{K}' とすると、左右それぞれ

$$[\bar{u}, \bar{v}, 1]^\top \propto \mathbf{K}[\bar{u}, \bar{v}, 1]^\top \quad (8)$$

$$[\bar{u}', \bar{v}', 1]^\top \propto \mathbf{K}'[\bar{u}', \bar{v}', 1]^\top \quad (9)$$

のように与えられる。

3.2 キャリブレーションの問題

以上の幾何モデルの下で次のようなキャリブレーションの問題を考える。空間の点 $X_{ij} (j = 1, \dots)$ を各視点 $i = 1, \dots$ で観測したときの座標を $\{(\tilde{u}_{ij}, \tilde{v}_{ij}) \mid (i, j) \in S\}$ 、右カメラについて $\{(\tilde{u}'_{ij}, \tilde{v}'_{ij}) \mid (i, j) \in S'\}$ と書く。ここで S は、視点 i の左カメラ画像上に空間の点 j が観測されているとき、 (i, j) を要素に持つような集合で、 S' は右カメラについて同様の集合である。これらを元に、左右のカメラ間の相対姿勢 \mathbf{T} を推定したい。ただしこの他に未知数として、各視点の姿勢 $\mathbf{T}_i (i = 1, \dots)$ と点の3次元位置 $\mathbf{X}_j (j = 1, \dots)$ があり、必要ならばこれも推定する。

ここでは、左右のカメラの内部パラメータおよびレンズ歪は、事前にキャリブレーションを行い既知であるとする。具体的には Zhang の方法 [12] を用いて行う。この結果、左カメラの画像上の点の座標 (\tilde{u}, \tilde{v}) を元に、 $[\bar{u}, \bar{v}, 1]^\top \propto \mathbf{K}^{-1}[\tilde{u}, \tilde{v}, 1]^\top$ を計算し、次に (7) 式を (u, v) について解くことで、歪と内部パラメータの影響を取り除いたカメラ座標 $[u, v, 1]^\top$ に正規化できる。なお後者は、(7) 式を歪修正量 $\Delta u = \bar{u} - u$ 、 $\Delta v = \bar{v} - v$ の式に書き換えた後、それらの式の右辺に u, v の推定値を代入し、 $\Delta u, \Delta v$ を求めて $u \leftarrow \bar{u} - \Delta u$ 、 $v \leftarrow \bar{v} - \Delta v$ と更新することを、 $u = \bar{u}$ 、 $v = \bar{v}$ を初期値に反復することで計算できる。そこで以下では、上の手順で正規化された両カメラの観測点 $\{(u_{ij}, v_{ij}) \mid (i, j) \in S\}$ および $\{(u'_{ij}, v'_{ij}) \mid (i, j) \in S'\}$ を元に、上述の未知パラメータを求めることを考える。

4. 方法

4.1 特徴点の追跡と左右の画像間対応付け

まず、対応点の集合 $\{(u_{ij}, v_{ij}) \mid (i, j) \in S\}$ および $\{(u'_{ij}, v'_{ij}) \mid (i, j) \in S'\}$ を得る。経時とともに得られる画像列に対し、各カメラで空間の同一点を追跡するとともに、左右のカメラ間でも同一点の対応関係を求める。この処理について述べる。

まず、左右のカメラ別々に特徴点の追跡を行う。 i 番目の視点で得た左右のカメラの画像をそれぞれ I_i, I'_i と書く。この処理では左右のカメラともに同じ処理を行うので、以下では左カメラのみについて説明する。

まず I_i から特徴点を抽出する。実験では SURF を用いた。 $i + 1$ 番目の画像 I_{i+1} に対しても同様に特徴点を抽出した後、特徴点の対応付けを行う ($I_i \leftrightarrow I_{i+1}$)。対応付けには各特徴点の特徴量 (128 個の実数からなるディスクリプタ) のユークリッド距離を用いる。各特徴点について、この距離が最小となるものを $I_i \rightarrow I_{i+1}$ および $I_{i+1} \rightarrow I_i$ の各方向について探す。このとき、視点間の車両およびカメラの移動量の小ささを前提に、特徴点の移動量は画像上で一定以下になると仮定し、各特徴点について、相手の画像上でその点の位置を中心とする入力画像サイズの 0.35 倍の矩形内を探す。こうやって得た両方向の対応のうち、双

方向に対応する点のペアのみを残す。

次にこうして得た対応点の組に対し、5点アルゴリズム [8] を用いた RANSAC を実行し、エピソード条件を用いた外れ値の除去を行う。また、5点アルゴリズムによって推定された基本行列 $E_{i,i+1}$ を分解した4通りの姿勢変化に対し、対応点の3次元位置を計算し、カメラ座標系の z 座標が正となるものの数が最大になる姿勢変化を記録し、同時にそれ以外対応点のペアを外れ値と認定して削除する。この2度のチェックを経て、連続する画像ペアの間で対応点のペアを得る。

これを繰り返すことで、空間の同一点を連続画像上で追跡した特徴点の軌跡が得られる。具体的には、 $I_i \leftrightarrow I_{i+1}$ のペアと $I_{i+1} \leftrightarrow I_{i+2}$ のペアが、 I_{i+1} の上で同じ特徴点を共有する場合、このペアをマージして、一つの空間の点の3枚の画像 (I_i, I_{i+1}, I_{i+2}) での軌跡とする。空間の同一点に対し、何かのはずみで2つの分離された軌跡を得る場合があり得るが、この場合は異なる2つの点の2つの軌跡であると見なす。

以上の処理は左右のカメラの画像 I_i, I'_i でそれぞれ独立に行う。次に、こうして得た左右それぞれの空間の同一点の軌跡に対し、対象とする空間の点と同じである点を選ぶ。選ばれた点の左右の画像上の軌跡は、これら画像間の対応を与えることになる。

そのために、各視点 i ごとに、左右の画像間 ($I_i \leftrightarrow I'_i$) で対応を求める。各特徴点について、相手の画像上でその点の位置を中心とする画像サイズの0.6倍の矩形内で、特徴量のユークリッド距離が最短となる特徴点を選び、左右双方向に対応するペアのみを対応候補とする。この処理は各視点ごとに行うため、左右の特徴点軌跡どうして見ると矛盾する場合が生じる(例えば i と $i+1$ とで、左画像の特徴点軌跡が対応する右画像の特徴点軌跡が異なるなど)。そこで、左の各特徴点軌跡について、右の特徴点軌跡のうちで対応候補を最も多く与えたものに対応付ける。これを右画像から左画像への向きでも同様に行い、双方向に対応付けられた特徴点軌跡を対応させる(すなわち空間の同じ点を左右の画像上で観測した軌跡と見なす)。

以上のようにして、空間の点ごとの特徴点軌跡を得る。これは3つに分類することで、左カメラおよび右カメラのどちらかのみを観測された点の軌跡と、両方のカメラから(1視点以上)観測された点の軌跡の3つである。次節に述べるバンドル調整では、これら軌跡をすべて使って最適化を行う。

4.2 パラメータの初期値の生成

推定すべき未知パラメータは、左右のカメラ間の相対姿勢 T および、各視点の姿勢 $T_i (i = 1, \dots)$ と点の3次元位置 $X_j (j = 1, \dots)$ である。

左右のカメラの相対姿勢 T は、空間の点のうち左右のカ

メラでともに観測されたものを選び、その画像間対応を元に基本行列を計算し、その分解不定性を点の奥行きが正となることを元に解決し、得る。

T_i は、上の特徴点追跡の過程で、連続2視点間で5点アルゴリズムを実行し基本行列 $E_{i,i+1}$ 、およびこれを正しく分解した回転と並進成分 $\delta R_i, \delta t_i$ を得ているので、基準座標をとった左カメラのそれを使って次のように求める ($i > 1$)。

$$T_i = \begin{bmatrix} \delta R_i & \delta t_i \\ 0^T & 1 \end{bmatrix} T_{i-1}. \quad (10)$$

X_j は、その軌跡の開始視点と終了視点を選び、上のようにして決まった両視点の姿勢を用いて、その2視点間でステレオ復元を行って決める。右カメラにしか観測されていない点を除き、左カメラの観測を優先して決める。いずれの場合でも、グローバルな基準座標系 ($i = 0$ の視点の左カメラのカメラ座標系) で表現し直して初期値とする。左カメラで観測された点は上のようにつめた T_i を用いて、右カメラの場合にはさらに T を用いて座標変換する。

4.3 ロバストバンドル調整

点 X_j の視点 i の左カメラの画像座標について、観測値(レンズ歪と内部パラメータの影響を除く正規化後のもの)を $[u_{ij}, v_{ij}]$ 、未知パラメータの推定値から計算されるその推定 $[\hat{u}_{ij}, \hat{v}_{ij}]$ 、と書くとき、両者の差

$$\delta u_{ij} = \hat{u}_{ij} - u_{ij} \quad (11)$$

$$\delta v_{ij} = \hat{v}_{ij} - v_{ij} \quad (12)$$

の2乗和 $\delta u_{ij}^2 + \delta v_{ij}^2$ が再投影誤差となる。同様に右カメラについても $(\delta u'_{ij}, \delta v'_{ij})$ を定義する。

次のような観測された点についての再投影誤差の総和

$$f = \frac{1}{2} \sum_{(i,j) \in S} \delta u_{ij}^2 + \delta v_{ij}^2 + \sum_{(i,j) \in S'} (\delta u'_{ij})^2 + (\delta v'_{ij})^2 \quad (13)$$

を、未知パラメータについて最小化するのが通常のバンドル調整である。ただし S および S' は、それぞれ左右カメラの観測点の集合とする。ロバストバンドル調整にはいくつかのやり方が考えられるが、ここではロバスト関数 ρ を用いて上のコストを次のように書き換える。

$$f = \frac{1}{2} \sum_{(i,j) \in S} \rho(\delta u_{ij}^2) + \rho(\delta v_{ij}^2) + \sum_{(i,j) \in S'} \rho((\delta u'_{ij})^2) + \rho((\delta v'_{ij})^2) \quad (14)$$

表記を単純化するため、 $(i,j) \in S \cap S'$ の各インデックスおよび u, v 座標を新しい通し番号 $i = 1, \dots$ で表すことにし、(14) 式を

$$f(\mathbf{p}) = \frac{1}{2} \sum_i \rho(e_i^2) = \frac{1}{2} \sum_i \rho_i \quad (15)$$

と書き換える ($\rho_i \equiv \rho(e_i^2)$)。ガウスニュートン法を導くた

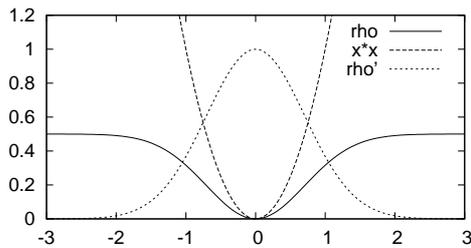


図 1 $\rho(x^2)$ と x^2 の比較．サンプルごとの重みを与える $\rho'(x^2)$ も同時に表示． $c = 1$ の場合．

め、 \mathbf{p} まわりに微小量 $\delta\mathbf{p}$ についてテーラー展開し

$$f(\mathbf{p} + \delta\mathbf{p}) \approx f(\mathbf{p}) + \sum_i \rho'_i e_i \left(\frac{de_i}{d\mathbf{p}} \right)^T \delta\mathbf{p} + \delta\mathbf{p}^T \left[\sum_i \left(\frac{de_i}{d\mathbf{p}} \right) (\rho'_i + \rho''_i e_i^2) \left(\frac{de_i}{d\mathbf{p}} \right)^T \right] \delta\mathbf{p} \quad (16)$$

を得る．ただし $\rho'_i = \rho'(e_i^2)$, $\rho''_i = \rho''(e_i^2)$ である．勾配ベクトルとヘッセ行列を

$$\mathbf{g} \equiv \sum_i \rho'_i e_i \left(\frac{de_i}{d\mathbf{p}} \right) \quad (17)$$

$$\mathbf{H} \equiv \sum_i (\rho'_i + \rho''_i e_i^2) \left(\frac{de_i}{d\mathbf{p}} \right) \left(\frac{de_i}{d\mathbf{p}} \right)^T \quad (18)$$

と定義する．ここではダンピング付（ガウス）ニュートン法を採用し，反復計算におけるパラメータの更新を

$$(\mathbf{H} + \lambda\mathbf{I})\delta\mathbf{p} = -\mathbf{g} \quad (19)$$

の解によって与える． λ は誤差が減少すれば 10 倍，増加すれば 1/10 する方法で制御した．

実験では ρ には Welsch の関数を用いた．すなわち

$$\rho(x; c) = \frac{c^2}{2} \left[1 - \exp\left(-\frac{x}{c^2}\right) \right] \quad (20)$$

である． c は最初大きな値を設定しておき，上の反復計算を実施して収束後， c を小さな値に切り替えて再度，上の反復計算を行う．これを 2, 3 度繰り返す．実験では，画像の横幅を長さ 1 としたとき， $c = 5, 0.05, 0.005$ の 3 段階に切り替えた．

4.4 コスト関数のロバスト化について

(14) 式に示したように，コスト関数は，特徴点の x, y 座標ごとの誤差の 2 乗の関数（例えば $\rho(\delta u_{ij}^2)$ ）としたが，これは外れ値であるかそうでないかの切り分けを特徴点の x, y 座標それぞれで行っていることに相当する．通常のロバストバンドル調整では，この切り分けを（少なくとも）画像上の観測点単位で行い，例えば $\rho(\delta u_{ij}^2 + \delta v_{ij}^2)$ のようにするのが普通である．その背景には，観測の誤りは特徴点の対応付けによるので，例えば y 座標のみが外れ値となるというような場合を考えにくいことがある．



図 2 左：実験システムの全体図．右：左カメラの様子．

それでも上のようにした理由は次のようなものである．第一に計算が簡単になること，第二に上のようにしても推定には悪影響を与えないと考えられることによる．

後者は具体的には次のようなことである．ある 3 次元点を復元可能な最小の観測は，2 枚の画像上に観測されている場合である，仮に点の空間座標のみが未知であると考えれば，4 つの観測に対し未知数は 3 つあることになるので，空間座標を定めた上でさらに 1 自由度分がカメラパラメータの推定に寄与することになる．この場合に，もし観測点単位で外れ値の判定を行ったとすると，2 枚のいずれかの画像上の観測が外れ値と判定されることで，実質的に観測が 1 枚の画像上にしかないということになり得る．そうすると，この点の空間座標は一意に定めることができず，全体の系は縮退することになる（もしロバスト関数の代わりに明示的な観測の除去を行っていれば，ヤコビ行列が特異となる）．上のように座標ごとに外れ値の切り分けを行うことにすれば，そのような場合でも未知数 3 つに対応する 3 つの観測は形式的にインライアとなり，空間座標を過不足なく定めることができる．その点の空間座標は誤って推定されているかもしれないが，カメラパラメータの推定には何ら影響を及ぼさない．今精度よく推定したいのはカメラパラメータのみであるので，このようにしても問題がない．

5. 実験結果

上述の方法の有効性，特に時間軸方向の情報を利用することの有効性とそれは外れ値除去に効果があることによるというわれわれの予想を検証するため，実験を行った．

5.1 実験の方法

図 2 のように，ワゴン車のルーフにカメラを 2 台取り付けした．カメラは Point Grey Research 社の Grasshopper2 を使用し，レンズは FUJINON の HF16HA-1B（画角約 $30 \times 23^\circ$ ）を用いた．画像は 640×480 画素とし，IEEE1394 経由で PC に毎秒 30 フレームのレートで取り込まれる．実験では 2 フレームおきに画像を取り出して用いた．つまり，100 視点の場合 200 フレーム（約 7 秒弱の期間）の画像を用いることになる．

左右のカメラ間の相対姿勢を推定する今の問題では，推定精度の検証それ自体も難しい．色々なやり方が考えられ

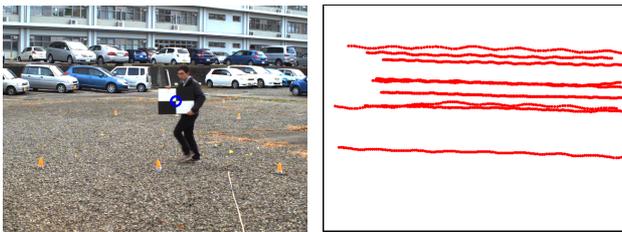


図 3 左：検証用の対応点取得の様子（右カメラ）. 右：取得した対応点の軌跡．

るが，ここではエピポラ条件を正確に満たすかどうかによって精度を評価することとした．つまり，左右のカメラの画像上で，専用のパターンを使って高精度に対応点

$$M \equiv \{(\tilde{u}_i, \tilde{v}_i), (\tilde{u}'_i, \tilde{v}'_i) \mid i = 1, \dots, n\} \quad (21)$$

を n 個求め，各対応点についてのエピポラ線からの距離を左右双方向に計算し，その平均をとった．つまり，左画像の点を指定した右画像上のエピポラ線およびその反対方向のエピポラ線

$$\mathbf{l}_i = \mathbf{E} \begin{bmatrix} \tilde{u}_i \\ \tilde{v}_i \\ 1 \end{bmatrix}, \quad \mathbf{l}'_i = \mathbf{E}^\top \begin{bmatrix} \tilde{u}'_i \\ \tilde{v}'_i \\ 1 \end{bmatrix} \quad (22)$$

を求め，それらについて

$$E_{epi} = \sqrt{\frac{1}{2n} \sum_{i=1}^n \text{dist}[(\tilde{u}'_i, \tilde{v}'_i), \mathbf{l}_i]^2 + \text{dist}[(\tilde{u}_i, \tilde{v}_i), \mathbf{l}'_i]^2} \quad (23)$$

を計算した．ただし $\text{dist}[\mathbf{p}, \mathbf{l}]$ は平面上の点 \mathbf{p} と直線 \mathbf{l} の距離である．また上の式の画像座標は歪を除去した後のものである．すなわち，観測された画像座標を \mathbf{K} を使ってカメラ座標に変換し，歪を除去した後，再度 \mathbf{K} を使って画像座標に直す．

エピポラ幾何を精度評価に用いたのは，それがステレオカメラの主な用途である密なマッチングの精度をまず最初に左右するからである．

実験ではこのような対応点を図 3 のようにチェッカーパターンを用いて求めた．また，奥行きを 3 段階に変え，またパターンを保持する高さも 3 段階に変えることで，点群が空間で適度な広がりを持つようにした．実験では，車両走行前と走行後に 2 セットこのような対応点組を得，それぞれ M_{before} , M_{after} と書く．これらは，車両走行に伴ってカメラの姿勢が変動していないことを確認するために利用する． M_{before} は 1691 組， M_{after} は 1807 組であった．

5.2 比較した方法

\mathbf{T} をセルフキャリブレーションのやり方で推定する方法はいくつか考えられる．性能比較の目的で，上述のロバストバンドル調整（ここでは Robust-BA と呼ぶ）の他に，ロバスト化していないバンドル調整（BA と呼ぶ）と，左右

表 1 キャリブレーション用画像群を用いた推定結果の E_{epi} (単位画素)．行がパラメータの推定に使ったシーケンスを，列が E_{epi} を算出するのに用いたシーケンスを表す．

	M_{before}	M_{after}
M_{before}	0.312	0.455
M_{after}	0.441	0.323

のカメラの画像間での点对応のみを使う方法（Two-view と呼ぶ）の 2 つを用いた．BA は (13) 式のコストを最小化し， \mathbf{T} を推定する．Two-view は，3 節の方法で取り出した特徴点軌跡のうち，左右のカメラの画像間に対応付けられたものを取り出し，これを用いて， \mathbf{T} を推定する方法である．この場合対応点は実質的に 2 台の固定カメラについてのもものと見なせ，2 視点の SfM をバンドル調整で実行し， \mathbf{T} を推定する．左右のカメラの相対姿勢を推定するのに利用できる情報は，(外れ値の存在を除けば) 両カメラ間の対応のみであり，Robust-BA, BA と Two-view でここは変わらないことに注意する．(つまり例えば Robust-BA が，左右カメラ間の対応点の異なる集合を使っているということはない．)

5.3 結果

まず M_{before} および M_{after} に対し，Two-view を用いてそれぞれ \mathbf{T} を推定した．推定した \mathbf{T} を， M_{before} および M_{after} を使って E_{epi} を計算した結果を表 1 に示す．表は行が \mathbf{T} を推定するのに使ったデータを，列が推定した \mathbf{T} の精度を評価するのに用いたデータを表す．推定と検証に同じデータを用いた場合でも誤差が 0.3 画素あるが，これはキャリブレーションパターン（図 3）の特徴点抽出誤差に由来すると考えられる．推定と検証に異なるデータを用いた場合の誤差（約 0.4 から 0.5 画素）は，走行前後 M_{before} , M_{after} でそれほど値が変わらないことから，走行中にカメラの姿勢が変化するようなことはなかったことが確認できる．さらにこの値（0.4-0.5 画素）が，セルフキャリブレーションが目標とすべき値と言える．

路上を走行した連続画像を対象に，Robust-BA, BA, Two-view それぞれを適用した．実験では 100 視点分の画像シーケンスを 26 個使った．うち 3 つのシーケンスのある視点における左右のカメラの特徴点追跡の様子および，Robust-BA による 3 次元復元結果を図 4 に示す．Robust-BA は大域最小解に収束したかどうかを判定するのは難しいが，3 次元復元の結果およびコスト関数の値から 26 シーケンスすべて同様に収束した． E_{epi} を用いた精度評価の結果を表 2 に示す．表では最も精度のよかったものを太字で表している．

結果は，26 シーケンス中 19 シーケンスで Robust-BA の誤差が最も小さかった．いくつかのシーケンスで，BA や Two-view の精度が最良となったのは，これらのシーケンスで外れ値の混入がたまたま少ないか，ほとんどなかった

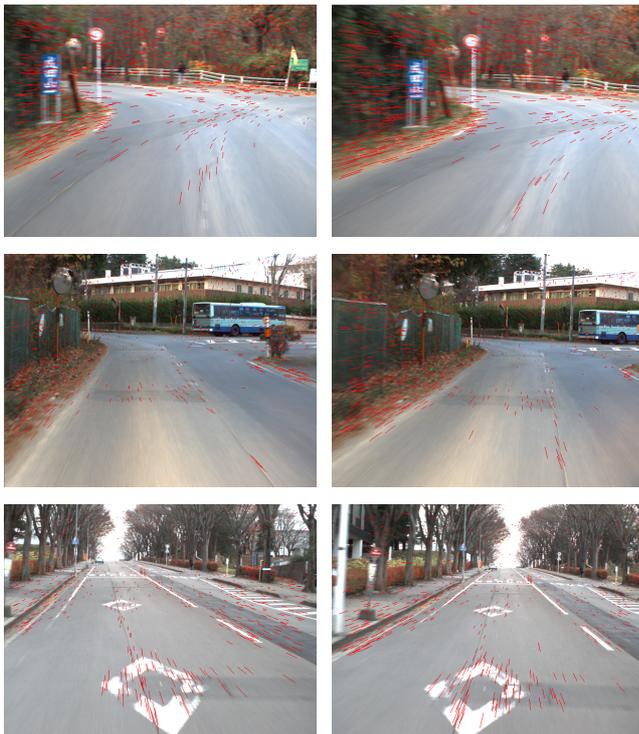


図 4 上から順に各列は c, f, z 各シーケンスのある視点での左右のカメラの画像と特徴点の追跡状況, および SfM 復元結果 (点群と 100 視点の位置). 左右は同じ復元を異なる視点から見たもの.

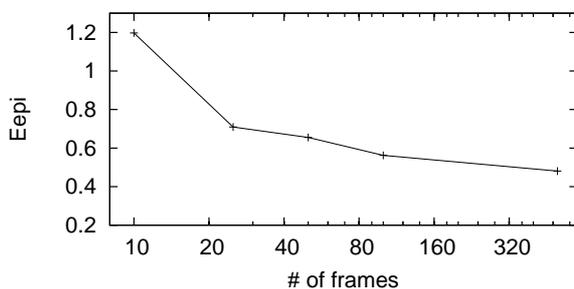
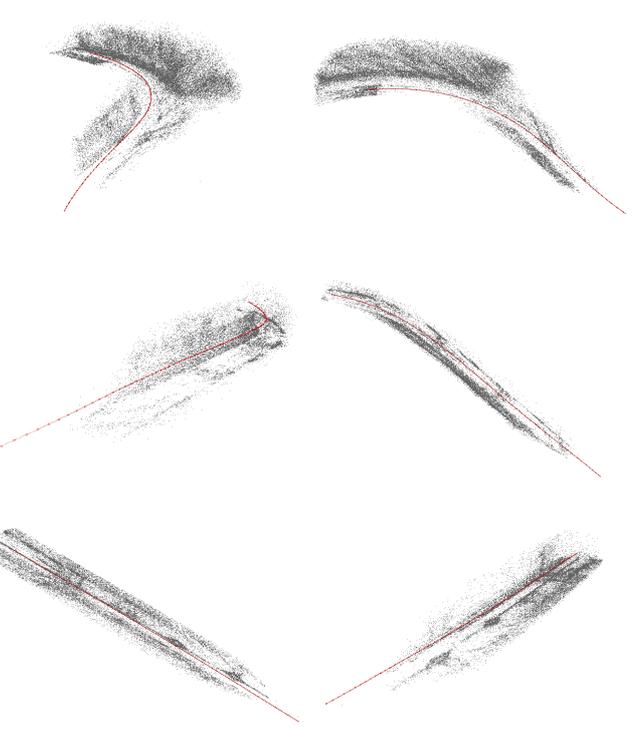


図 5 使用した視点数と E_{epi} の変化.

ことによるものと考えられる. BA や Two-view の結果で, 良くなる場合はあるものの時に非常に大きな誤差を返していることはこのことを裏付ける. これに対し Robust-BA の値は全シーケンスで安定しており, その値は 0.5 画素前後である. 上述のように, キャリブレーションパターンを用いたオフライン推定の結果に若干劣る程度と言え, その利便性を考えれば十分実用的であると言える.

また, 視点の数を 10 から 500 まで変化させた時の E_{epi} の変化を図 5 に示す. 視点の数を増やすほど順調に精度が向上していることが分かる. ただし今の場合, E_{epi} の値そのものが誤差を含む M_{before} の対応点を用いて計算されているため, 誤差はどこかで飽和するはずである. (ただし理論的には, 推定精度は視点数に比例して向上し続けると考えられる.)



6. まとめ

車載カメラのセルフキャリブレーションの方法を示し, 実験評価を行った. 意図した通り, 車載カメラから得られる画像の時間軸方向の対応付けを用いて, ロバストバンドル調整に基づく全体最適化を行うことで, 推定精度を向上させられることが分かった.

参考文献

- [1] Azarbayejani, A. and Pentland, A. P.: Recursive Estimation of Motion, Structure, and Focal Length, *IEEE Trans. PAMI*, Vol. 17, pp. 562–575 (online), DOI: <http://dx.doi.org/10.1109/34.387503> (1995).
- [2] Bailey, T., Nieto, J., Guivant, J. E., Stevens, M. and Nebot, E. M.: Consistency of the EKF-SLAM Algorithm., *Proc. IROS*, IEEE, pp. 3562–3568 (2006).
- [3] Broida, T. J., Chandrashekar, S. and Chellappa, R.: Recursive 3-D motion estimation from a monocular image sequence, *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 26, No. 4, pp. 639–656 (1990).
- [4] Dang, T., Hoffmann, C. and Stiller, C.: Continuous Stereo Self-Calibration by Camera Parameter Tracking, *IEEE Trans. Image Processing*, Vol. 18, No. 7, pp. 1536–1550 (2009).
- [5] Davison, A. J.: Real-Time Simultaneous Localisation and Mapping with a Single Camera, *Proc. ICCV*, pp. 1403– (online), available from (<http://portal.acm.org/citation.cfm?id=946247.946734>) (2003).
- [6] Hartley, R. and Zisserman, A.: *Multi-View Geometry in*

表 2 27 種類の異なる道路を走行した 100 視点分の画像列に対する推定結果。左右の表はそれぞれ M_{before} , M_{after} で検証したもの。100 視点分の SfM をロバストバンドル調整したもの (Robust-BA), 同様にバンドル調整したもの (BA), 左右の画像間のみの対応を使ってバンドル調整したもの (Two-View) の結果。点数 (左右カメラ共有点) とは SfM の結果得られた空間中の点数 (うち左右両カメラで観測された点数)。一点あたりの観測数とは空間中の点数を得られた全ての観測数で割ったもの。

	点数 (共有点)	観測/点	Robust-BA	BA	Two-view		Robust-BA	BA	Two-view
a	58335(3973)	3.06	0.727	0.831	0.599	a	0.750	0.961	0.664
b	63963(3917)	3.35	0.402	0.758	0.699	b	0.585	0.851	0.814
c	55713(3322)	3.00	0.471	0.779	1.59	c	0.598	0.931	1.76
d	72235(3610)	2.90	0.667	0.605	0.954	d	0.684	0.681	1.02
e	70785(3885)	3.06	0.532	0.735	0.489	e	0.585	0.730	0.542
f	61225(4045)	3.27	0.502	0.845	1.36	f	0.558	0.787	1.30
g	69775(3194)	3.14	0.530	1.09	1.18	g	0.701	1.19	1.13
h	61163(2194)	2.70	0.559	0.855	0.955	h	0.581	0.935	1.01
i	80936(2698)	2.83	0.543	0.994	1.00	i	0.550	0.912	1.07
j	74587(3911)	3.05	0.598	1.78	1.53	j	0.656	1.89	1.47
k	70033(2843)	2.72	0.456	0.746	1.04	k	0.595	0.707	1.18
l	66890(2671)	2.91	0.833	1.73	0.863	l	0.976	1.84	0.930
m	72644(3352)	3.02	0.623	0.738	0.697	m	0.665	0.686	0.789
n	68765(3889)	3.07	0.507	0.742	0.479	n	0.607	0.692	0.542
o	71097(4022)	3.21	0.579	0.410	0.590	o	0.694	0.525	0.671
p	76273(4476)	3.20	0.520	1.01	0.553	p	0.574	1.16	0.607
q	72755(3371)	3.01	0.483	0.778	0.716	q	0.583	0.827	0.778
r	69758(3802)	3.18	0.489	1.35	0.691	r	0.574	1.27	0.769
s	63782(3066)	2.93	0.658	0.874	0.683	s	0.717	0.925	0.742
t	62374(3473)	2.95	0.458	0.668	0.535	t	0.527	0.652	0.543
u	60846(3814)	3.11	0.604	0.497	0.553	u	0.693	0.535	0.609
v	64453(3672)	2.98	0.470	0.724	0.517	v	0.537	0.683	0.511
w	61807(3475)	2.96	0.533	0.997	1.01	w	0.655	0.932	1.03
x	79764(3208)	3.01	0.439	0.896	0.727	x	0.526	0.851	0.815
y	61734(2793)	2.91	0.653	0.475	0.729	y	0.675	0.576	0.730
z	71209(3440)	3.16	0.425	0.737	0.476	z	0.549	0.865	0.578

- Computer Vision*, Cambridge University Press (2000).
- [7] Maybank, S. J.: Filter Based Estimates of Depth, *Proc. BMVC*, pp. 349–354 (1990).
- [8] Nistér, D.: An Efficient Solution to the Five-Point Relative Pose Problem, *IEEE Trans. PAMI*, Vol. 26, pp. 756–777 (online), DOI: 10.1109/TPAMI.2004.17 (2004).
- [9] Strasdat, H., Montiel, J. M. M. and Davison, J.: Real-time Monocular SLAM: Why Filter?, *Proc. ICRA* (2010).
- [10] Sturm, P.: Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction, *Proc. CVPR* (1997).
- [11] Wiles, C. and Brady, M.: Ground Plane Motion Camera Models, *Proc. ECCV* (1996).
- [12] Zhang, Z.: A flexible new technique for camera calibration, *IEEE Trans. PAMI*, Vol. 22, No. 11, pp. 1330–1334 (2000).
- [13] Broggi, A., Bertozzi, M. and Fascioli, A.: Self-calibration of a stereo vision system for automotive applications, *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, Vol. 4, pp. 3698 – 3703 vol.4, DOI: 10.1109/ROBOT.2001.933193 (2001).
- [14] Nedeveschi, S., Vancea, C., Marita, T. and Graf, T.: Online Extrinsic Parameters Calibration for Stereovision Systems Used in Far-Range Detection Vehicle Applications, *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 8, No. 4, pp. 651 –660, DOI: 10.1109/TITS.2007.908576 (2007).
- [15] Zhang, Z., Luong, Q.-T. and Faugeras, O.: Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, *Robotics and Automation, IEEE Transactions on*, Vol. 12, No. 1, pp. 103 –113, DOI: 10.1109/70.481754 (1996).