

多視点 2 眼一体型カメラのキャリブレーション Multiple Binocular Stereo Camera Calibration

多田 和広[†]
Kazuhiro Tada

池畑 諭[†]
Satoshi Ikehata

相澤 清晴^{† ‡}
Kiyoharu Aizawa

1. はじめに

多視点から撮影した画像群を用いる際、画像の対応関係から各カメラの位置関係を推定する取り組みが以前より行われている [1] [2]。これまでの研究で、単眼カメラを用いた様々な手法の検討が行われてきた。しかし、それらの手法で精度の高い推定結果を得るには数多くの画像が必要であった。

本論文では、撮影デバイスとして 1 台で 2 眼を有する 2 眼一体型カメラを用い、カメラ間の既知の関係を利用することで、より正確なカメラキャリブレーションを行うことを目的とする。まず、単眼カメラで行われてきた従来の解決法について述べる。次に、2 眼一体型カメラを用いた場合に、2 眼一体型カメラで同時に撮影した 2 枚の画像間において領域ベースの対応付けを利用する方法を提案する。最後に、これらを利用した実験を行い、その結果を示す。

2. Structure from Motion

Structure from Motion (SfM) は、同一のシーンを複数の視点から撮影した画像群を用いて、そのシーンの 3 次元形状とカメラの位置を同時に復元する方法である。実際は、複数の画像に共通して写る点の座標群 z_{ij} から、未知パラメータであるカメラパラメータ p_i と 3 次元空間上の点の座標 q_j を、バンドルアジャストメントの考え方をを用いて求める [3]。観測座標 $z_{ij} = [u_{ij}, v_{ij}]^T$ と、推定中のパラメータ p_i, q_j によって計算される座標 $\bar{z}_{ij} = [\bar{u}(p_i, q_j), \bar{v}(p_i, q_j)]^T$ の 2 乗距離の和

$$E(p_1, \dots, p_m, q_1, \dots, q_n) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n \{(u_{ij} - \bar{u}(p_i, q_j))^2 + (v_{ij} - \bar{v}(p_i, q_j))^2\} \quad (1)$$

が最小になるよう、 p_i と q_j を求める。通常、最小 2 乗アルゴリズムを用いて反復的に計算する。

SfM を用いたカメラ位置の推定を行うソフトウェアの 1 つとして Bundler が挙げられる [4]。あるシーンを様々な位置から撮影した画像を用意するだけで実行できるため、広く利用されている。Bundler は、計算に考慮するカメラを 1 つずつ加えていくことで、逐次的にカメラ位置を推定する。まず、これまでに得られている 3 次元点を最も多く観測するカメラを計算に追加する。次に追加したカメラに対して DLT 法でカメラパラメータを推定する。最後に、このカメラパラメータを初期値として、全体のカメラ位置をバンドル調整で最適化する。これを、カメラが追加できなくなるまで繰り返す。

このように、Bundler が担当するのは、特徴点群の最適な対応付けを求めるためのバンドル調整と因子分解である。その前処理となる特徴点抽出については別途行う。一般的な特徴点として、回転やスケール変化に強い SIFT [5] を用いることが多い。

3. 領域ベースのステレオマッチング

このように SfM では、画像間の対応付けに特徴点ベースのマッチングを行うことが多い。特徴点によるマッチングはある程度大きな画像の変化に強い。しかし、特徴点ベースのマッチングには、特徴点群が疎になりやすいという問題がある。本研究の場合、2 眼一体型カメラで同時に撮影した 2 枚の画像間においては、その他の画像ペア間よりも多くの情報が抽出できると考えられるため、このような疎なマッチングだけでは十分な対応情報が得られないと考えられる。

そこで本論文では、領域に基づいたマッチングを加えることを提案する。領域ベースマッチングとは、基準点を中心とした局所ブロック画像と、入力画像の局所ブロック画像を一定の尺度を用いてマッチングを行う手法である。指定した基準点に対する対応点を探索することができるため、密な基準点を設定することで、画像全体に分布する密な対応点を得ることができる。特徴点マッチングと比べ、大きな画像変形に対応できないことが問題であるが、ステレオペアの関係にある画像間では大きな画像変化はないため、十分に利用できる手法であると言える。

本手法では、領域ベースによるステレオマッチングで広く用いられている SAD を対応付けの尺度として用いる [6]。テンプレートの大きさを $N_1 \times N_2$ 、テンプレートを $T(i, j)$ 、対象画像を $I(i, j)$ としたとき、SAD は以下の式によって計算される。

$$R_{SAD} = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} |I(i, j) - T(i, j)| \quad (2)$$

探索ウィンドウに対してテンプレートを動かしながら SAD を計算し、SAD の値が最も小さくなった位置を対応点とする。

4. 提案手法

本論文では、従来の画像の対応付けに加え、2 眼一体型カメラで同時に撮影した 2 枚の画像間に対して領域ベースマッチングの手法による対応付けを行う手法を提案する。本手法により、2 眼一体型カメラのステレオペアに対してより多くの対応情報を抽出でき、精度の高いカメラキャリブレーションができると期待される。

提案手法のフローチャートを図 1 に示す。2 眼一体型カメラで撮影した画像群から任意の 2 枚を選び、SIFT による特徴点の対応を抽出する。もし選出した 2 枚の画像が 2 眼一体型カメラで同時に撮影した画像の場合、同時に SAD による点の対応付けを抽出する。SAD による対応を計算する際、GrabCut [7] によって撮影対象を画像から切り抜いたマスクを用意し、このマスクの範囲内で点の対応を探索する。SAD 対応点の基準点は 5 ピクセル間隔で取る。これらの特徴点の対応付けを全ての画像ペアに対して行う。最後に、この対応付けを用いて、SfM によるカメラ位置の推定を行う。本実

[†]東京大学大学院 情報理工学系研究科

[‡]東京大学大学院 情報学環・学際情報学府

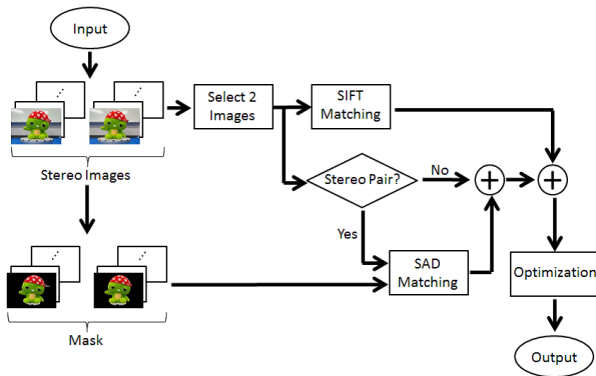


図 1: フローチャート

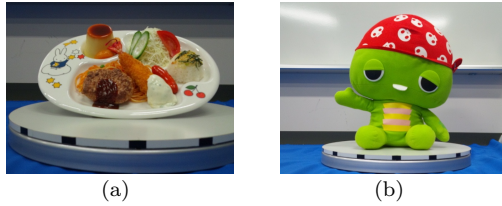


図 2: 使用した画像

験では、このカメラ位置の推定処理を Bundler が行う。

5. 実験

5.1. 実験の概要

実験に用いた撮影対象は図 2 の 2 種類であり、対象物体をターンテーブルに乗せ、一定間隔で回転させながら撮影したものである。7 視点 14 枚の画像を撮影し、解像度は 640x480 で、ターンテーブルの回転角の間隔は 30° である。また、撮影地点からターンテーブルの中心までの距離は (a) が 50cm, (b) が 80cm である。

これらの画像を用いて、図 1 の手法でカメラ位置を推定し、各カメラ行列から相対的なカメラ間の変換行列を求め、カメラ間の位置関係を調査する。正解値として、Zhang らの手法 [8] によって計算されたカメラ間の変換行列を用意する。また従来法として、SIFT のみで画像の対応を取り、Bundler で最適化したデータを用意する。これら 3 種類のデータの結果を比較する。

5.2. 結果

表 1 に、図 2(a) においてステレオペアの関係にある 2 画像間の相対的なカメラ行列の一例を示す。この表から、従来手法よりも提案手法の方が正解値に近い値を出していることがわかる。図 2(a) の場合、回転行列では平均して 57.7%、並進ベクトルでは 14.7%、各値が正解値に近づいている。同様に (b) では回転行列が 9.6%、並進ベクトルが 3.4% 正解値に近づいている。

表 2 は、図 2(a) においてステレオペアの関係にある 2 画像間の相対的なカメラ行列の各値の分散を示したものである。これを見ると、提案手法の方が従来手法より分散値が高い傾向にあることがわかる。そのため、提案手法の方がパラメータの安定性は低いと考えられる。これは提案手法がより多くの特徴点を抽出していることと、ピクセル単位のマッチングを行っているために、全体最適化のステップで誤った対応を取ってしまう可能性が増したためと考えられる。

6. まとめ

本論文では、2 眼一体型カメラを用いて多視点から撮影した画像を用いたキャリブレーション法について、

表 1: ステレオペア間の相対カメラ行列の一例

	回転行列			並進ベクトル
正解値	0.999513	-0.000866	0.031209	-1.610782
	0.000939	0.999997	-0.002352	-0.014132
	-0.031206	0.002380	0.999510	0.023949
従来手法	0.999582	0.014305	0.025140	-0.206485
	-0.014309	0.999898	-0.000028	0.005516
	-0.025138	-0.000332	0.999684	-0.043211
提案手法	0.999593	0.003743	0.032846	-1.485325
	0.003702	0.999910	-0.002431	0.031472
	-0.031445	0.002365	0.999673	0.014057

表 2: ステレオペア間の相対カメラ行列の分散

	回転行列			並進ベクトル
従来手法	6.32025E-09	4.22955E-05	2.54722E-06	-0.002410
	4.21461E-05	2.35225E-09	2.07648E-06	1.97543E-06
	1.61798E-06	1.85504E-06	8.28100E-09	0.001339
提案手法	0.052618	0.031240	0.132737	0.455846
	0.039978	0.024934	0.091544	0.362380
	0.124801	0.100347	0.049722	0.006139

ステレオペア間における対応点の拡張を利用した手法を提案した。実験の結果、カメラ間の相対的な位置関係を従来手法よりも正確に求められるケースが多い。一方で、求められる値の分散値が高く、安定した結果は得られにくいこともわかった。

今後は、対応点抽出後のカメラ位置の最適化手法について検討を進める予定である。本実験では、最適化ステップを従来手法に一任していたが、ステレオペアとそれ以外の画像のペアで使用する特徴量が異なる以上、最適化手法においてもその差異を利用した工夫が必要である。一例として、ステレオペア間ではカメラ間の位置関係が既知であるため、ユークリッド変換の不定性を除却した最適化手法を導入し、それ以外の画像のペア間の最適化と分割する手法を考えている。

参考文献

- [1] K. Haming et al., "The structure-from-motion reconstruction pipeline - A survey with focus on short image sequences", *Kybernetika*, 2010
- [2] 池畑 諭 ら, "多視点ステレオのための信頼度に基づく高精度デプスマップの推定と融合", *MIRU*, 2011
- [3] B. Triggs et al., "Bundle Adjustment - A Modern Synthesis", *Lecture Notes in Computer Science*, Vol.1883, 2000
- [4] N. Snavely et al., "Modeling the World from Internet Photo Collections", *IJCV*, 2007
- [5] D. G. Lowe, "Object recognition from local scale-invariant features", *ICCV*, 1999
- [6] K. Ambrosch et al., "SAD-based stereo matching using FPGAs", *Embedded Computer Vision*, 2009
- [7] C. Rother et al., "GrabCut" - Interactive Foreground Extraction using Iterated Graph Cuts", *ACM Trans. Graph*, 2004
- [8] Z. Zhang et al., "A flexible new technique for camera calibration", *IEEE TPAMI*, 2000