2R-1

モーションステレオによる大域空間推定

坂本 優太† 久保田 光一‡

中央大学大学院 理工学研究科 情報工学専攻†‡

要約: 単眼カメラによる 3 次元復元手法に単眼カメラを用いたモーションステレオ法がある．この手法は単眼カメラ 1 台のみを使用するため手軽に利用することができるが, センサやステレオカメラを利用する手法と比較すると精度が低く, 非局所空間を対象とする場合には誤差は無視できないものとなってしまう．そこで, 本研究では単眼カメラを用いたモーションステレオ法による非局所空間の 3 次元復元システムを作成するにあたり, 特徴点群の 3 次元座標とカメラの運動パラメータの最適化を試み, 実装および実験により有用性を調査した．

キーワード: 3 次元復元, モーションステレオ

1 はじめに

3 次元復元手法は古くから研究されており, 最近では技術の発達により産業やエンターテインメントなどさまざまな分野での需要があり, 更なる技術の発展が望まれているため現在でも研究が盛んに行われている．近年での 3 次元復元研究で用いられている手法は主に赤外線センサ等のセンサ類を用いる手法やステレオカメラを用いる手法, 3 台以上のカメラを用いる多眼ステレオ手法, 単眼カメラを用いたモーションステレオ法などがある．この他にも複数ある手法の中から, 本研究では単眼カメラを用いたモーションステレオ法に着目した．単眼カメラはデジタルカメラや web カメラ, モバイル端末など多くの身近にあるデバイスが利用可能であり, デバイス固有の問題が発生しにくい利点がある．

しかし, 単眼カメラを用いたモーションステレオ法は他の手法と比べて復元精度が低いと言われている．その精度が局所範囲での復元では許容できるとしても, 非局所空間の推定を目的とした場合, 精度が低いと誤差が積もることで大きなズレが生じ正確な推定が難しくなる．

本研究では, 単眼カメラを用いたモーションステレオ法による非局所空間の 3 次元復元システムを作成するにあたり, 動画から検出される特徴点群の 3 次元座標と動画を撮影したカメラの運動パラメータの最適化を試み実装した．さらに, 実験によりその有用性を調査した．

ここで, 本研究で大域空間や非局所空間と表しているのは, 画像 1 フレームに収まらない範囲と定義している．

2 類似研究

単眼カメラによる動画や画像を使用した研究に [1] や [2] などがある．[1] では特徴点の計測の一部にトータルステーションを用いてカメラ位置推定の効率化と高精度化を行なっている．また, [2] では特徴点の対応付けに [3] の技術を利用している．最近ではリアルタイムでの高精度 3 次元復元を可能にした DTAM[4] の論文が発表され, 注目を集めている．

3 実装

本研究で実装したプログラムにおける処理の流れは次の順で行われる．

1. 動画の入力
2. 特徴点抽出と対応付け
3. 特徴点の 3 次元座標, カメラの運動パラメータの計算
4. 統合処理
5. 推定モデル

さらに, 各ステップでの説明を以下に記載する．

3.1 動画の入力

使用する動画は, 焦点距離を一定にして撮影した wmv 形式の動画とする．

3.2 特徴点抽出と対応付け

まず, 特徴点抽出の前処理として, 安定した特徴点を求めることができるようにガウシアンフィルタで画像を平滑化し, グレースケール変換を行った．ガウシアンフィルタは σ 値が 10 で 11×11 のカーネルを使用した．

特徴点抽出には疎な特徴集合を求めることができる Shi-Tomasi 法を利用し, 2 画像間における特徴点の対応付けは画像ピラミッドを利用して Lucas-Kanade 法を反復実行することにより, 疎な特徴集合に対するオプティカルフローで対応付けを行った．Lucas-Kanade 法で最小二乗法により類似性を求める際に用いる行列の固有値が特徴点の運動量計算のしやすさを示しているため, Shi-Tomasi 法を利用することで画像中から大きな固有値を持つ特徴点を検出でき, 精度の高い特徴点の対応付けが可能になる．

3.3 特徴点の 3 次元座標, カメラの運動パラメータの計算

この処理は Tomasi・Kanade の因子分解法 [5] を利用することとし, 因子分解法による 3 次元復元を実装したプログラム*を利用した．

このプログラムでは使用する全てのフレームに同一特徴点群が存在し対応付けしていなければならない．しかし, 本研究では非局所空間を対象としているため全てのフレームに同一特徴点が存在することはない．したがって, 本研究では図 1 のように動画全体を複数範囲に分割して各範囲ごとに特徴点抽出と対応付けをすることで範囲内で同一特徴点群の存在を可能にし因子分解法を適用した．また, 各範囲での因子分解法はフレーム数を変化させて複数回行い以下に記載した統合処理により最適値を求めた．ここで, 処理するフレームを変化して解を複数回求めたのは, 入力動画の撮影時に速い移動や光などの障害によりフレームによっては特徴点が正確に求まらないことがあり, コンピュータでは全ての特徴点を正確に対応付けすることは難しく誤対応が発生してしまい, 特徴点群の 3 次元座標やカメラの運動パラメータに大きな誤差が生じてしまう可能性があるからである．この処理を全範囲で行うことで入力動画全体を網羅した．

3.4 統合処理

本研究では前ステップで分割して求めた複数の特徴点群の 3 次元座標, カメラの運動パラメータ値の最適値を求める処理を統合処理と呼ぶことにする．

一範囲内における統合処理では, 微小時間ではカメラの移動

Nonlocal Area Space Recognition with Motion Stereo Vision

† Yuta SAKAMOTO, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

‡ Koichi KUBOTA, Information and System Engineering Course, Graduate School of Science and Engineering, CHUO University

* 3次元復元アルゴリズムに関する公開プログラム, "http://www.suri.cs.okayama-u.ac.jp/program-3d.html"

Copyright ©2012 Information Processing Society of Japan.
All Rights Reserved.

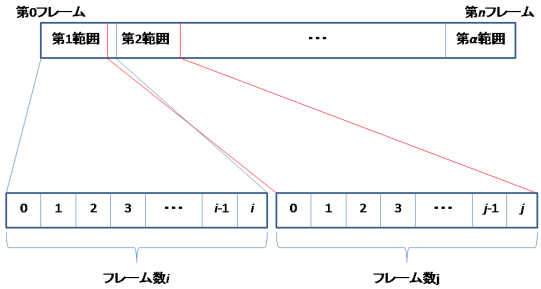


図 1　処理するフレームの選択

は直線であると仮定し, 前ステップで利用したフレーム数ごとにカメラの移動経路を直線で近似して最も直線に近いデータ群を一範囲の解候補とし, フレームごとにカメラ位置の 3 次元座標の重心をそのフレームでのカメラ位置とした. また, 特徴点群の 3 次元座標においても各特徴点の重心を解とした.

この各範囲での統合処理で得た結果を合わせることにより動画全体における特徴点群の 3 次元座標値とカメラの運動パラメータを求めることとした.

3.5　推定モデルの出力

出力では第 1 フレームのカメラ位置を原点とした各特徴点の 3 次元座標値とカメラの運動パラメータを出力した.

4　実験

この節では実装したプログラムが一範囲内で特徴点が正確に求まっていないフレームや対応付けが正確に行われなかったフレームの影響を減らし最適なカメラの移動経路を求めているかを実験を通して調査する. なお, 以下では特徴点群の重心を原点とした上方を X 軸, 水平方向を Y 軸, 原点から第 1 フレームの画像平面への垂線方向を Z 軸とし単位をピクセルとした xyz 左手座標系に関して記述してある.

実験データは第 1 フレームと最終フレームが図 2 となるように, Y 軸の負方向への平行移動をして撮影した 13 フレームの動画を使用した. また, 第 1 フレームにて抽出した特徴点がそれ以降のフレームの画像上のピクセルにどの様に対応付けられているかを確認するために, 実験データから抽出した特徴点群の追跡結果を図 3 に示した. 全てのフレームで対応付けした特徴点を画像座標系の図としてプロットし, 第 1 フレームの特徴点を×印で示した. 図 3 からは X 方向に変化している部分が一部見受けられるため, 対応付け時に誤対応が発生していることがわかるが概ねの対応付けは正確に行われている.

表 1(a) は一範囲内の全フレームを使用して求めた各フレームでのカメラ位置座標を表し, 表 1(b) は一範囲内で統合処理を行い最適値と判断された各フレームでのカメラ位置座標を表している. この最適値は連続する 4 フレームを使用して統合処理で求められた値となった.



図 2　実験データの第 1 フレームと最終フレーム

因子分解法では全ての特徴点の対応が正しく求まっていなければならないが, コンピュータで計算する際には全ての対応を正しく求めるのは困難であり図 3 のように誤った対応付け

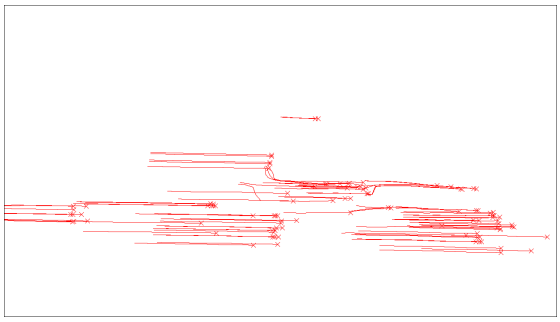


図 3 第 1 フレームの画像上の特徴点として抽出されたピクセルとそれ以降のフレーム上の対応する特徴点のピクセルとを線分で結んだ追跡結果 (特徴点 100 個)

表 1　推定されたカメラの位置座標

	(a)			(b)		
フレーム番号	X 座標	Y 座標	Z 座標	X 座標	Y 座標	Z 座標
1	482	660	2594	480	657	2582
2	482	576	2603	472	624	2580
3	482	496	2608	448	617	2583
4	482	408	2615	455	617	2583
5	480	313	2620	456	594	2581
6	476	213	2627	457	590	2575
7	457	71	2521	451	582	2580
8	459	-9	2540	451	495	2569
9	460	-84	2549	435	430	2566
10	454	-146	2550	471	212	2550
11	454	-210	2548	469	137	2549
12	454	-284	2554	467	52	2549
13	451	-339	2545	466	-12	2549

を行なってしまうことがしばしばある. その影響を受けて全フレームを使用した表 1(a) では Z 軸方向へ大きな誤差が生じてしまっているが, 少数のフレームから求めて統合し最適値と判断された場合は誤った特徴点対応が局所範囲のみにしか影響を及ぼさないので, 結果に及ぼす影響は全てのフレームを使用した場合よりも少なくなっており, 直線的に動いている経路を求めることができていることがわかる.

しかし, Y 軸方向のみの動きでは全フレームを使用した表 1(a) の方が最もらしい値が求まっている.

5　まとめ

実験の結果からでは表 1(a) と (b) のどちらが良いとは断定することができなかったが, 小節 3.3 の処理で使用するフレーム数の違いによる結果で大きな差が発生することがわかった.

今後はこの結果と統合処理による動画全体での違いにも考慮して最適化を行なっていく必要がある.

参考文献

[1] 佐藤 智和, 神原 誠之, 横矢　直和, 竹村　治雄, “マルチベースライン ステレオ法を利用した動画像からの屋外環境の三次元モデル化”, バーチャルリアリティ学会論文誌, No.7, Vol.2, 2002.

[2] 山田健人, 金澤靖, 金谷健一, 菅谷保之, “2 画像からの 3 次元復元の最新アルゴリズム”, 情報処理学会研究報告, 2009-CVIM-168-15 (2009-8), pp. 1-8.

[3] 金澤靖, 金谷健一, “2 画像間の特徴点対応の自動探索–シーンに関する知識を上手に使う”, 日本工業出版, 画像ラボ 15(11), 20-23, 2004-11.

[4] Richard A. Newcombe, Steven J. Lovegrove and Andrew J. Davison, “DTAM: Dense Tracking and Mapping in Real-Time”, ICCV2011.

[5] 金谷健一, 菅谷保之, “因子分解法の完全レシピ”, 電子情報通信学会技術報告 PRMU2003-118 (2003-10), pp.19-24.

Copyright ©2012 Information Processing Society of Japan.
All Rights Reserved.