

NAIST-IS-MT9951049

修士論文

単眼動画画像からのマーカと自然特徴点の 自動追跡による三次元復元

佐藤 智和

2000年 2月 9日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

佐藤 智和

審査委員： 横矢 直和 教授
千原 國宏 教授
竹村 治雄 助教授

単眼動画像からのマーカと自然特徴点の 自動追跡による三次元復元*

佐藤 智和

内容梗概

動画像からの三次元復元は、物体認識、ロボットナビゲーション、複合現実感など、様々な分野への応用が可能である。しかし、特徴点の出現や消失がおこる動画像や、長時間撮影された動画像からの三次元復元は従来、精度や計算量の点で問題があった。そこで本論文では、三元位置関係と色・形状が既知の複数個の基準マーカを利用し、基準マーカと自然特徴点を画像上で自動追跡することによって、高速に精度良く三次元復元を行う手法を提案する。提案手法では、透視投影カメラモデルの仮定のもとで、カメラの外部パラメータと自然特徴点の三次元位置を毎フレームで推定することにより、基準マーカが観測できないフレームにおいても、高速かつ精度の良いカメラの外部パラメータの推定が可能である。加えて、自然特徴点の追跡においてロバスト推定を用いることで自然特徴点の誤追跡を抑え、精度の高い三次元復元を行う。また、処理の最後に動画像全体での最適化を行うことにより、逐次処理による推定誤差の蓄積を最小化し、更なる復元精度の向上を図る。実験では、屋内および屋外の実環境を撮影した動画像を入力とした三次元復元を行い、提案手法の有効性を示す。

キーワード

三次元復元, 単眼動画像, 透視投影カメラモデル, マーカ追跡, 自然特徴点追跡

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9951049, 2000年2月9日.

3-D Reconstruction

from a Monocular Image Sequence

by Tracking Markers and Natural Features*

Tomokazu Sato

Abstract

Three-dimensional (3-D) reconstruction from an image sequence has been widely used for object recognition, robot navigation, mixed reality, and so on. However, there are some problems concerning calculation cost and the precision in the 3-D reconstruction from a long sequence of images that may contain features come in and out. In this paper, we propose a new 3-D reconstruction method from a monocular image sequence, which uses a number of markers whose positions in real world, color and shape are known, as well as natural features. In this method, the camera parameters and 3-D positions of natural features are estimated efficiently in every frame by tracking these markers and natural features automatically. Furthermore, we employ a robust estimation approach for tracking natural features to prevent mistracking. Finally, the accumulation of estimation errors is minimized by optimization over the whole input. To show the feasibility of the proposed method, we demonstrate experimental results of 3-D reconstruction from real image sequences of both indoor and outdoor scenes.

Keywords:

3-D reconstruction, monocular image sequence, perspective camera model, marker tracking, natural feature tracking

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9951049, February 9, 2000.

目次

1. はじめに	1
2. 三次元復元に関する従来研究と本研究の位置付け	3
2.1 受動的三次元復元の従来手法	3
2.2 動画像を用いた三次元復元	6
2.2.1 動画像のみによる復元	6
2.2.2 事前知識を利用した復元	8
2.3 本研究の位置付けと方針	8
3. 基準マーカと自然特徴点の追跡による三次元復元	10
3.1 三次元復元手法の概要	10
3.2 カメラパラメータの定義	12
3.3 基準マーカの追跡	14
3.4 自然特徴点の追跡	14
3.4.1 Harrisのインタレストオペレータを用いた自然特徴点の候補位置の検出	15
3.4.2 ロバスト推定を用いた暫定カメラパラメータの推定	17
3.4.3 探索範囲を限定したマッチングによる自然特徴点の追跡	18
3.5 特徴点の追跡結果を用いたカメラパラメータの推定	18
3.6 自然特徴点の三次元位置の推定	21
3.7 特徴点の信頼度の算出	23
3.8 自然特徴点の追加と削除	24
3.9 動画像系列全体での最適化	25
4. 実験	26
4.1 屋内環境における実験	26
4.2 屋外環境における実験	34
4.3 考察	34

5. まとめ	41
謝辞	42
参考文献	43

図目次

1	三次元復元の手法と復元対象の制約および復元に必要な情報 . . .	4
2	提案手法の処理の流れ	11
3	カメラ座標系と世界座標系の関係	13
4	追跡の容易な自然特徴点の例	16
5	入力画像と Harris オペレータによる特徴量の出力結果	16
6	推定される自然特徴点の三次元位置	22
7	屋内環境での実験における入力画像	29
8	屋内環境での実験における特徴点の追跡結果	30
9	屋内環境での実験における MatchMove	31
10	カメラの位置・姿勢と自然特徴点の三次元位置および面の構成結果 (屋内環境の場合)	32
11	動画全体での最適化による累積誤差の推移(屋内環境の場合) . .	33
12	屋外環境での実験における入力画像	35
13	屋外環境での実験における自然特徴点の追跡結果	36
14	屋外環境での実験における MatchMove	37
15	カメラの位置・姿勢と自然特徴点の三次元位置および面の構成結果 (屋外環境の場合)	38
16	動画全体での最適化による累積誤差の推移(屋外環境の場合) . .	39

表目次

1	受動的な三次元復元手法	4
2	実験環境	26
3	実験に用いたパラメータ	27

1. はじめに

画像処理による三次元形状の復元は、物体認識、ナビゲーション、複合現実感などさまざまな分野への応用が可能である。そのため、古くはステレオ法に代表されるように、画像処理技術を用いた三次元計測の手法が研究されてきた [1]。二枚または三枚の画像からのステレオ法においては、カメラ位置および姿勢を既知として、エピポラ幾何を用いて画像上での探索範囲を直線に限定し、対応点の探索を行う。このため、厳密なカメラキャリブレーションが必要となるが、屋外環境のような、カメラ間の距離（ベースライン）を大きく取る必要がある環境においては、厳密なカメラキャリブレーションを行うことは困難である。そこで、多数の画像を用いて画像間の対応点を複数与えることにより、カメラのキャリブレーションを行う self-calibration の研究や、ビデオカメラを用いて対象を撮影し、動画像を入力とすることで三次元形状とカメラパラメータを同時に推定する shape from motion へと発展している。

中でも、動画像を入力とする三次元復元の手法においては画像上の特徴点を自動追跡することができるため、画像間の特徴点の対応づけが自動化できる。このため、環境の三次元復元を自動化するための手法として動画像を入力とする三次元復元手法の研究が盛んである [2, 3, 4, 5, 6]。しかし、特徴点の出現や消失がおこる動画像や、長時間撮影された動画像からの三次元復元においては、精度や計算量の点で問題があるため、大規模かつ複雑な環境の復元が必要とされる場合には、現在でも手動による復元を行うことがほとんどであり、これは多大な労力を要する。

そこで本論文では、広域な環境を高速に精度良く復元することを目的とし、三元位置関係と色・形状が既知の複数個の基準マーカを利用し、基準マーカと自然特徴点を画像上で自動追跡することによって、三次元復元を行う手法を提案する。提案手法では、透視投影カメラモデルの仮定のもとで、カメラの外部パラメータと自然特徴点の三次元位置を毎フレームで推定することにより、基準マーカが観測できないフレームにおいても、精度の良いカメラの外部パラメータの推定が可能である。加えて、自然特徴点の追跡においてロバスト推定を用いることで自然特徴点の誤追跡を抑え、精度の高い三次元復元を行う。また、処理の最後に動画像全体での最適化を行うことにより、逐次処理による推定誤差の蓄積を最小化し、更なる

復元精度の向上を図る。ただし、提案手法ではカメラの内部パラメータは既知であるとする。また、一般的には三次元位置関係が既知の基準マーカから線形演算によってカメラパラメータを推定するためには、6点以上の基準マーカが必要である [7] ため、入力動画像の初期フレームにおいては基準マーカが6点以上指定される必要がある。

以下、2章では、三次元復元に関する従来研究と、本研究の位置付けについて述べる。3章では、特徴点として基準マーカと自然特徴点を用い、撮影時のカメラパラメータと自然特徴点の三次元位置を復元する手法について述べる。4章では、実際に現実環境を撮影した動画像を入力として実験を行い、本手法の特徴と有用性について考察する。最後に5章で、本論文のまとめと今後の課題を述べる。

2. 三次元復元に関する従来研究と本研究の位置付け

対象に接触することなく三次元復元を行う手法は、レーザーレンジファインダやより一般的なパターン光投影による測定のように、撮影対象に積極的に働きかけることで対象の形状を計測する能動的な三次元復元手法と、撮影対象への働きかけは行わず、画像処理のみにより復元を行おうとする受動的な三次元復元手法に大別される。能動的な三次元復元手法では、比較的精度良く対象の形状を測定することができるが、屋外環境のような広域な環境の測定に用いることは困難であり、また装置が複雑で高価であるという問題があるため、本論文では扱わない。

本章では受動的な三次元復元の従来手法について述べ、なかでも、図1に示すように、復元対象に関する制約条件と事前に必要とされる情報が少ない、動画像を用いる三次元復元手法について詳述する。

2.1 受動的三次元復元の従来手法

画像処理による受動的な三次元復元手法は、表1に示すように、陰影による手法 (Shape from Shading)、レンズ焦点法 (Shape from Focus)、輪郭による手法 (Shape from Contour)、ステレオ法、動画像による手法 (Shape from Motion)、など多くの方法がある。図1に示すように、それぞれ復元対象に関する制約条件や既知とする情報量が異なる。どの手法も、現在までに多くの研究が成されているが、任意の撮影対象を制約条件無しに自動で復元できるような手法は無いといえる。そこで、以下では各手法に関して特徴と問題点を述べる。

陰影による手法

陰影による手法 (Shape from Shading) は、撮影対象の陰影画像と反射特性、光源情報を利用して三次元復元を行う手法である。光源情報を既知として一枚の陰影画像から復元を行う手法 [8] や、photometric stereo と呼ばれる複数の陰影画像から復元を行う手法 [9, 10] があるが、復元対象に関する滑らかさの制約条件などを用いることで、比較的精度の良い復元を行うことが可能である。しかし、屋外環境のように、対象の反射特性や照明条件が複雑な環

表 1 受動的な三次元復元手法

陰影による手法 (Shape from Shading)	画像により得られる陰影情報と対象の反射特性, 光源情報による復元手法
レンズ焦点法 (Shape from Focus)	レンズの焦点深度を変化させた複数の画像を用いた復元手法
輪郭による手法 (Shape from Contour)	複数の校正済みカメラにより撮影された画像の輪郭情報による復元手法
ステレオ法	複数の校正済みカメラにより撮影された画像の視差による復元手法
動画像による手法 (Shape from Motion)	移動カメラにより撮影された動画像上の特徴点の移動軌跡による復元手法

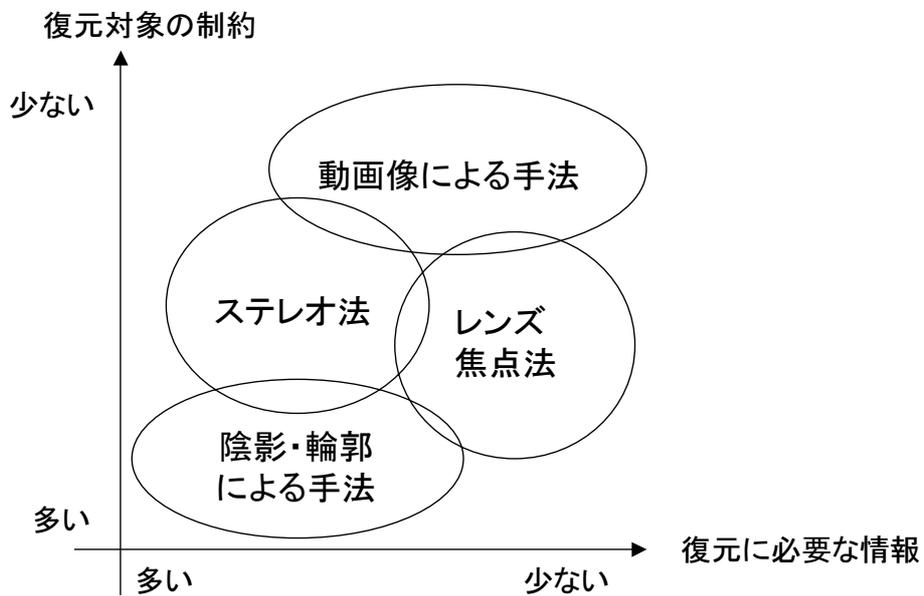


図 1 三次元復元の手法と復元対象の制約および復元に必要な情報

境に利用することは困難であり、また、必要な事前情報である光源情報や物体の反射特性を得ることが難しいといえる。

レンズ焦点法

レンズ焦点法 (Shape from Focus) は、レンズの焦点深度を変化させた複数の画像において、被写界深度の範囲外に映る対象に画像上のボケが発生するという光学的な特徴を利用し、奥行きを求める手法である [11, 12, 13]。また、異なるアプローチとして、焦点距離の異なる 2 枚の画像から画像上の全点での奥行きを計測する Shape from Defocus と呼ばれる手法が研究されている [14, 15]。これらの手法は、ともに必要とされる事前情報が少なく、比較的ロバストに奥行き情報を得ることができる。しかし屋外環境のように、撮影対象に対し数 m 以上カメラが離れた場合、十分に浅い焦点深度を設定することが難しく、満足な計測精度は期待できない。

輪郭による手法

輪郭による手法 (Shape from Contour) は、対象の輪郭情報を用いる手法である。撮影時のカメラ位置を既知とすることで、ロバストで復元精度の良い形状復元を行うことができる。最も基本的な手法は、回転台の上に撮影対象を配置し回転させ、固定カメラからその対象の輪郭を観測することにより対象の表面形状の復元を行うものである [16]。近年、ジャイロやマーカを用いることにより、任意のカメラ位置から対象を撮影する手法も提案されている [17, 18]。しかし、対象の輪郭抽出手法の検討が課題となり、一般的には、背景を単色とすることで対象の輪郭を抽出する手法が用いられる。そのため、一般的な環境にこの手法を適用することは困難である。

ステレオ法

ステレオ法は、位置の異なる複数のカメラを用いて対象を撮影し、三角測量の原理を用いて三次元復元を行うものである。最も基本的な手法である二眼ステレオ法では、カメラ位置・姿勢が既知の二台のカメラにより撮影された画像に対し、画像上で対応点を決定することによって、三角測量の原理により対象の奥行きを測定するものである。ステレオ法においては、正確な対

応点の探索が課題となっており、それを実現するために、動的計画法を用いる手法 [19] や、多重スケールを用いた手法 [20] など様々な手法が提案されている。しかし、完全といえる方法はなく、いずれも対応点のあいまいさや計算コストなどの問題が残されている。そこで、三台以上のカメラを用いる三眼ステレオ法が提案され [21, 22], 対応点のあいまいさを減少させる試みが成されているが、カメラの厳密なキャリブレーションが必要とされ、屋外環境のように、カメラ間の距離 (ベースライン) を長く取る必要がある環境において利用することは困難である。

動画像による手法

動画像を用い、撮影対象の特徴点を画像上で追跡することによりカメラパラメータと特徴点の三次元位置を復元する手法は Shape from Motion と呼ばれる。この手法は、カメラパラメータと特徴点の三次元位置の二つを復元する必要があるが、比較的容易に多くの対応点を自動で与えることができ、精度のよい復元が期待できる。CCD カメラの普及や計算機の高性能化とともに盛んに研究されるようになった [2, 23, 3, 5]。さらに、より広域環境の復元を目的とし、動画像に加えて、ジャイロや GPS などのセンサや事前知識を利用する手法が提案されている [4, 24, 25]。

上述のような各手法の特徴より、手軽に広域な環境を自動で復元することを目的とした場合には、動画像による手法が適しているといえる。そこで次節では、動画像を用いた復元手法について述べる。

2.2 動画像を用いた三次元復元

2.2.1 動画像のみによる復元

動画像のみを用いた復元手法の代表的なものに、カメラモデルを線形近似し、線形解法によって撮影対象の形状とカメラの運動パラメータを同時に求める因子分解法 [2] がある。この手法は対象形状・カメラ運動が自由であること、線形解法により高速かつ安定に解を得ることが可能であること、などの理由により、広く用い

られ、また手法に関する様々な拡張が試みられている [23, 26, 27, 28, 29, 30]. しかし問題として、以下の2点が挙げられる.

- (a) カメラモデルを線形近似するために、奥行きの高い対象の復元には大きな誤差が伴う
- (b) 動画像の全てのフレームに撮影されている特徴点を復元対象としてしているため特徴点が連続して観測できない動画像における形状復元は困難である

(a)の問題に対し、カメラモデルとして透視投影モデルを適用する試みがある [26, 27, 28]. しかし、撮影された画像間のエピポラ拘束を記述するための基礎行列や、非線形最適化が必要となるため、因子分解法が持つ、処理が単純で高速であるという利点が損なわれているといえる. また (b)の問題に対し、Polemanら [23] は、計測行列の空白部分を補間する手法を提案している. これは、数フレーム間の隠れ点には対応可能であるが、特徴点が画像外に移動してしまった場合など、長期間特徴点が観測できない場合には、計測行列の空白部分が増加し、大きな誤差として復元結果に表れる. このように、現在においても因子分解法における、(a)、(b)の問題は解消されていない.

一方、カメラモデルとして透視投影を仮定し、非線形最適化の手法によって画像上の観測誤差を最小化することで三次元復元を行う手法が提案されている [3, 5, 6, 31]. これらは透視投影モデルを仮定するため、奥行き方向に大きい対象を扱うことができ、また特徴点の出現や消失にも対応可能である. Szeliskiら [6] は、推定される特徴点の三次元位置を画像上に投影した位置と、実際に画像上で観測される特徴点の位置の距離の二乗和を誤差として定義し、これを誤差関数として非線形最適化を行うことで三次元復元を行う手法を提案した. しかし、局所解の問題や計算時間の問題など、実用面で多くの問題を残している.

また、Pollefeysら [5] は、まず射影的な三次元復元を行い、これを初期値として非線形最適化を行うことで三次元復元を行う手法を提案している. Sawhneyら [3] は、動画像を複数の部分に分け、各部に対して画像上の誤差を最小化し、統合処理を行った後に動画像全体で最適化を行うことで、三次元復元を行う手法を提案している. しかし、どちらの手法も動画像全体に対して非線形最適化を行うため、多

くの計算時間が必要とされ、特別な仮定がない限り大量の画像を扱うことは困難である。また、局所解の問題も充分解決されているとは言えない。

このように、動画像のみを用いた三次元復元では、復元精度や処理速度の面で問題があり、数百枚以上の画像を用いる広域な環境の三次元復元は困難である。

2.2.2 事前知識を利用した復元

動画像に加えて事前知識としてカメラの位置・姿勢や対象の形状に関する情報を利用して広範囲の復元を行う手法がある。上原ら [4] は動画像に加えて数値地図とGPSを利用する復元手法を提案し、広域な都市環境の三次元復元を試みている。戎野ら [24] は12台のカメラと2つの六角推ミラーを用いた高分解能パノラマステレオセンサを開発し、センサが等速直線運動するという仮定のもとで街並みの復元を行った。向井ら [25] は、ジャイロセンサを利用し、球面状の投影面を持つカメラモデルを採用することで、線形演算により三次元復元を行う手法を試みている。

一般的に、復元対象を限定せず、手軽に精度の良い復元を行うためには、利用する情報が容易かつ正確に得られることが重要である。しかし、上述の手法においては、センサ類の複雑さや、その計測誤差、また撮影時のカメラの動きの制約によって、復元対象が限定されたり、復元結果に大きな誤差が伴うという問題がある。

2.3 本研究の位置付けと方針

上述のような特徴により、復元対象を限定せず、手軽に広域の復元を行うためには、動画像による復元が適している。しかし、従来手法において、動画像のみによる手法は、計算時間や復元精度に問題があり、また、事前知識を用いる手法では、利用する情報が複雑で計測誤差を伴うために対象が限定されるという問題があった。

そこで、本論文では、広域な環境を高速に精度良く復元することを目的とし、単眼動画像に加えて、色・形状が既知の複数個の基準マーカを利用し、これらの三次元位置関係と初期フレームの画像上の座標を事前知識として用いる手法を提案する。このような事前知識を得ることは容易であり、かつカメラの動きや復元対象の制約も少ない。

提案手法では、基準マーカと自然特徴点を画像上での出現・消失を考慮しながら

ら自動追跡することによって、カメラの外部パラメータと自然特徴点の三次元位置を毎フレームで推定する。これにより、基準マーカが観測できないフレームにおいても、高速に精度の良いカメラの外部パラメータの推定が可能である。加えて、自然特徴点の追跡においてロバスト推定を用いることで自然特徴点の誤追跡を抑え、精度の高い三次元復元を行う。また、処理の最後に動画像全体での最適化を行うことにより、逐次処理による推定誤差の蓄積を最小化し、更なる復元精度の向上を図る。

ただし、提案手法では、カメラの内部パラメータおよび基準マーカの色や形状は既知であるとし、その情報を用いて画像上での追跡を行う。また、一般的には空間中の位置関係が既知の基準マーカから線形演算によってカメラパラメータを推定するためには、6点以上の基準マーカが必要である [7] ため、入力動画像の初期フレームにおいては基準マーカが6点以上指定されるものとする。

3. 基準マーカと自然特徴点の追跡による三次元復元

3.1 三次元復元手法の概要

本論文では、カメラの内部パラメータ、基準マーカの三次元位置関係および色・形状、初期フレームにおける基準マーカの画像上の座標を事前知識として用い、特徴点(基準マーカと自然特徴点)を毎フレームで追跡することにより、撮影時のカメラの外部パラメータと自然特徴点の三次元位置を推定する。ただし、撮影対象は静止した剛体であるものとする。

提案手法では図 2 に示すように、まず、初期フレームにおいて画像上で 6 個以上の基準マーカを指定する。指定された基準マーカに対して任意に設定された世界座標系における三次元位置関係と、その画像上の座標により、3.5 節に述べる手法を用いて初期フレームのカメラパラメータを推定する。次に、フレーム毎に特徴点を追跡しカメラパラメータを逐次推定する(図中 A)。この逐次処理における第 f フレームの処理の流れは、以下のようになる。

(i) 基準マーカの追跡

色・形状の情報を用い、基準マーカを追跡する

(ii) 自然特徴点の追跡

ロバスト推定を用い、第 $(f - 1)$ フレームの入力画像中に存在する自然特徴点を追跡する。

(iii) 追跡結果を用いたカメラパラメータの推定

追跡された特徴点の現フレームでの画像上の座標と、その特徴点の第 $(f - 1)$ フレームで推定された三次元座標から、特徴点の信頼度を考慮してカメラパラメータを推定する。

(iv) 自然特徴点の三次元位置の推定

第 f フレームまでに推定されているカメラパラメータ系列と追跡されている自然特徴点の画像上の座標系列から、自然特徴点の三次元位置を推定し、第 $(f - 1)$ フレームにおいて推定されていた三次元位置を更新する。

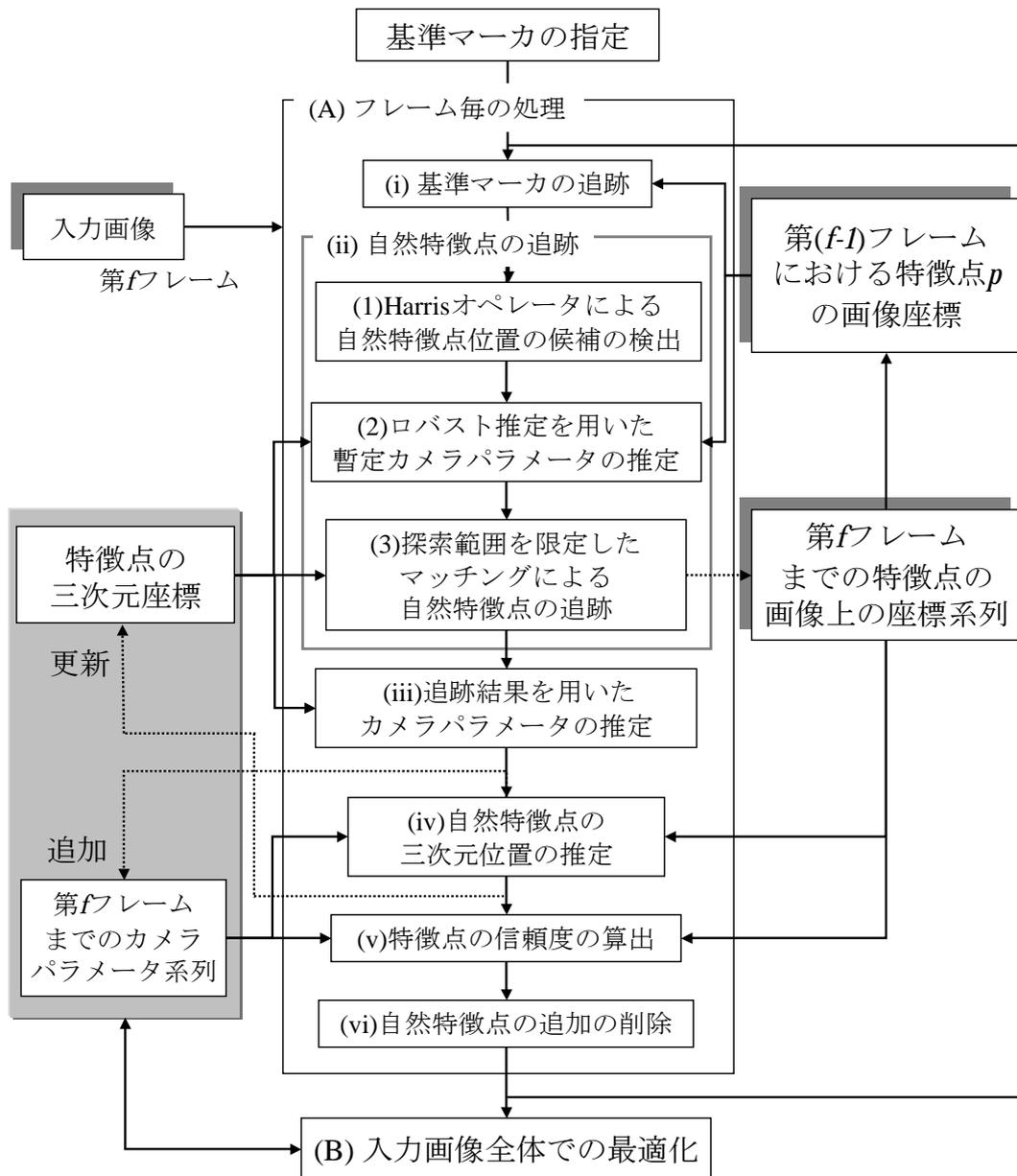


図 2 提案手法の処理の流れ

(v) 自然特徴点の信頼度の算出

自然特徴点の追加・削除の尺度や, 第 $(f + 1)$ フレームにおけるカメラパラメータ推定の重みとして利用するための自然特徴点の信頼度を算出する.

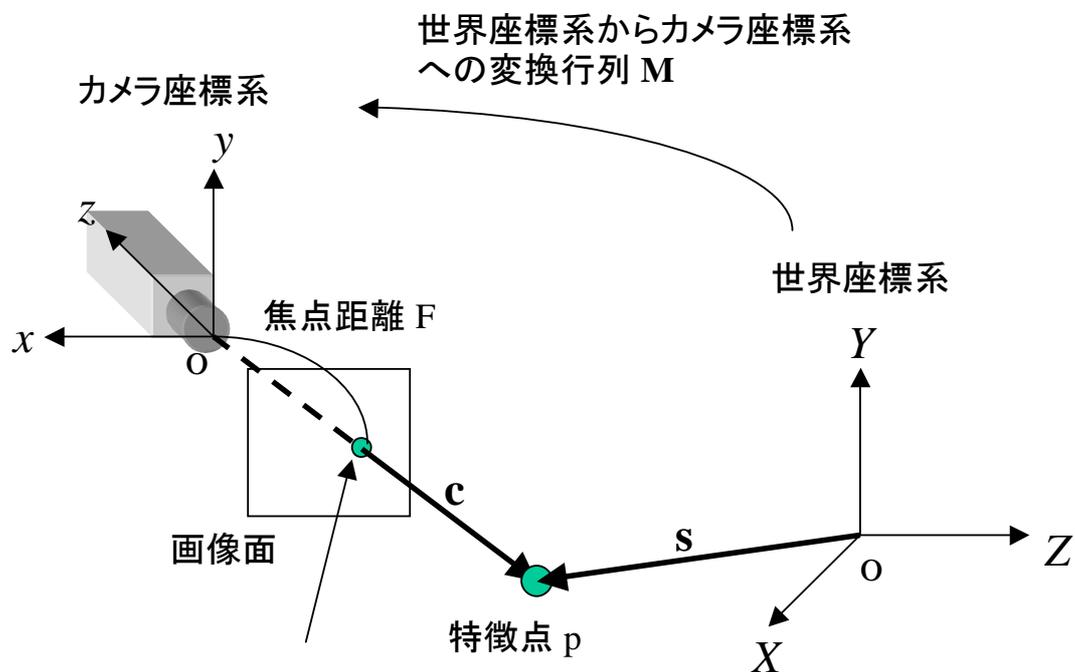
(vi) 自然特徴点の追加と削除

信頼度やマッチング誤差の尺度を用いて, 自然特徴点の追加と削除を行う.

最後に, 入力動画画像系列全体での最適化処理を行う (図中 B). この処理においては, 図中 A の逐次処理において推定された全てのフレームにおけるカメラパラメータと自然特徴点の三次元位置を初期値とし, 入力動画画像系列全体に対して, 画像上で観測される特徴点の位置と特徴点の三次元位置を画像上に投影した座標の二乗誤差 (再投影の誤差) を信頼度による重みつきで最小化する. これは, 非線形最適化問題となるが, 図中 A の逐次処理における推定結果が真値に近ければ, 非線形最適化問題において問題とされる局所解の問題を避け, また少ない繰り返し回数で大域最小解を得ることが可能となる. 以下, まず, 本論文で使用するカメラパラメータの定義を述べ, 続いて, 各処理について順に詳述する.

3.2 カメラパラメータの定義

本論文では, カメラの内部パラメータである, 焦点距離 F , CCD の画素のサイズ ccd_u, ccd_v , および, レンズ歪み率 k_1, k_2 は Tsai の手法 [32] などを用いて, あらかじめ計測しておくものとし, 外部パラメータであるカメラの位置・姿勢を推定する. 図 3 に示すように, カメラの外部パラメータは実世界に設定されている世界座標系から, カメラ座標系への変換行列 M で表すことができる. 本論文では, カメラの姿勢パラメータとして, カメラ座標系の 3 軸に対する回転角 (θ, δ, γ) と, 平行移動成分である $t = (t_1, t_2, t_3)^T$ の 6 つのパラメータを用いて, カメラパラメータ M を以下のように定義する.



特徴点 p の画像面への
投影座標 $\mathbf{x} = (u, v)$

図 3 カメラ座標系と世界座標系の関係

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

$$= \begin{pmatrix} c_1 c_3 + s_1 s_2 s_3 & s_1 c_2 & c_1 s_3 + s_1 s_2 c_3 & t_1 \\ -s_1 c_3 + c_1 s_2 s_3 & c_1 c_2 & s_1 s_3 + c_1 s_2 c_3 & t_2 \\ c_2 s_3 & -s_2 & c_2 c_3 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2)$$

ただし,

$$\begin{aligned} s_1 &= \sin \theta, & s_2 &= \sin \delta, & s_3 &= \sin \gamma \\ c_1 &= \cos \theta, & c_2 &= \cos \delta, & c_3 &= \cos \gamma \end{aligned} \quad (3)$$

また, 図3に示すように, ある点 p の世界座標系における三次元座標を $\mathbf{s} = (X, Y, Z)^T$, 点 p のカメラ座標系における三次元座標を $\mathbf{c} = (x, y, z)^T$ とすると, 世界座標系とカメラ座標系の変換は以下のように記述される.

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4)$$

レンズ歪みを考慮しなければ, 点 p の \mathbf{c} の画像上への投影座標 $\mathbf{x} = (u, v)$ は, 以下のように定義される.

$$(u, v) = \left(-\frac{F}{ccd_u} \frac{x}{z}, -\frac{F}{ccd_v} \frac{y}{z} \right) \quad (5)$$

3.3 基準マーカの追跡

提案手法においては, 既知である基準マーカの色・形状を用いた追跡を行う. ここでは, 基準マーカとして球状の単色マーカを用いる場合の追跡手法について述べる.

球状の単色マーカを用いる場合, どの位置から撮影してもマーカの画像上の重心位置と現実世界におけるマーカの三次元位置での重心が一致する. よって第2フレーム以降におけるマーカの追跡位置は, 第 f フレームにおいて, 第 $(f - 1)$ フレームのマーカの位置を中心とする一定幅の探索窓内で, 既知のマーカの色に近い色を持つ領域を抽出し, その面積が最大となる領域の重心位置を算出することで決定する.

3.4 自然特徴点の追跡

カメラの姿勢が大きく変化するような動画像では, 画像上での特徴点の見え方が大きく変化するために, 一般的に用いられる画像上の輝度勾配を用いた手法 [33]

や初期テンプレートを用いた自然特徴点の追跡手法 [34] では、以下の二つの問題があり精度の高い自然特徴点の追跡が困難である。

- (a) 見え方の変化によるテンプレートの中心位置のずれ
- (b) 周辺の類似パターンとの誤対応

提案手法では、(a) を解決するために、入力画像上のエッジのコーナなどの追跡の容易な点の検出に Harris のオペレータ [35] を用い、この出力が極大となる点を自然特徴点の候補位置とする。また、(b) を解決するために、自然特徴点の仮の追跡を行い、ロバスト推定を用いて暫定的なカメラパラメータを推定することで、追跡すべき自然特徴点の探索範囲を限定する。

本手法における、自然特徴点の追跡処理の手順は以下ようになる。

- (1) Harris のインタレストオペレータによる自然特徴点の候補位置の検出
- (2) ロバスト推定を用いた暫定カメラパラメータの推定
- (3) 探索範囲を限定したマッチングによる自然特徴点の追跡

以下に、これらの処理について順に詳述する。

3.4.1 Harris のインタレストオペレータを用いた自然特徴点の候補位置の検出

本研究では、追跡の際のテンプレートの中心位置のずれを抑えるため、Harris のインタレストオペレータ [35] の出力が極大値を取る座標を自然特徴点の候補位置として利用する。Harris のインタレストオペレータは、例えば図 4 の正方形窓で示すような、エッジのコーナなどの、追跡が容易な点で出力値 (特徴量) が大きくなるオペレータである。Schmid ら [36] の論文によれば、このオペレータは、回転、拡大などの画像の変形操作を行った場合に、他のインタレストオペレータに比べて、同一の点に特徴点が発見される再現度が最も高いオペレータであるとされている。

Harris オペレータによる入力画像上の座標 (x, y) の特徴量 H_{xy} は、まずガウシアンオペレータによる入力画像の平滑化処理を行い、次に一定の大きさの正方形窓 W において、画像上の輝度 I の勾配 I_u, I_v を用いて以下に示す行列 A を算出し、その行列の固有値 λ_1, λ_2 の極小値を求めることで得られる。

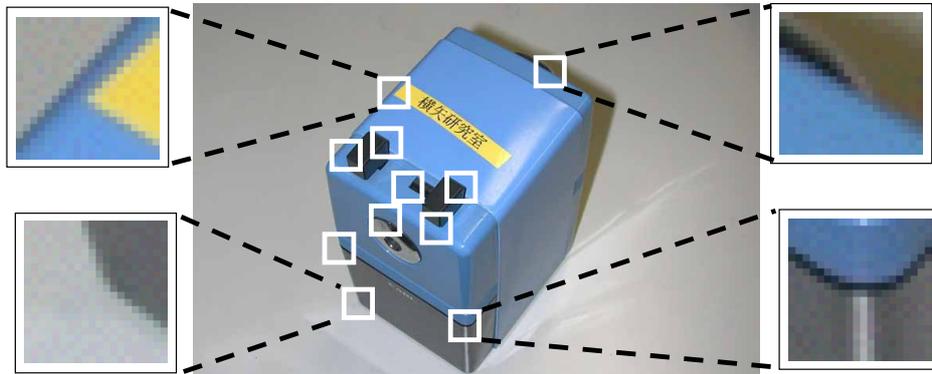
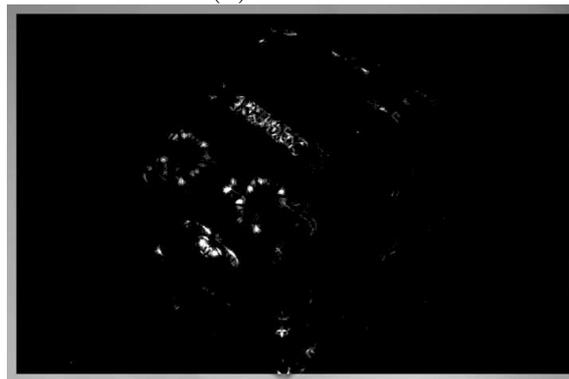


図 4 追跡の容易な自然特徴点の例



(a) 入力画像



(b) 特徴量を輝度値に変換した画像

図 5 入力画像と Harris オペレータによる特徴量の出力結果

$$\mathbf{A} = \sum_{\mathbf{x} \in \mathbf{W}} \begin{pmatrix} I_u(\mathbf{x})^2 & I_u(\mathbf{x})I_v(\mathbf{x}) \\ I_u(\mathbf{x})I_v(\mathbf{x}) & I_v(\mathbf{x})^2 \end{pmatrix} \quad (6)$$

$$H_{xy} = \min(\lambda_1, \lambda_2) \quad (7)$$

図5(a)に対して実際に Harris オペレータを用いて算出された特徴量を輝度値に変換した画像を同図(b)に示す. このように, 特徴量 H_{xy} の値が大きい点は, エッジのコーナなど, 周辺の領域に対し比較的追跡が容易な特徴点であると考えられるため, 本研究では, 特徴量 H_{xy} が極大値を取る位置を自然特徴点の候補位置 $\mathbf{c} = (c_0, c_1, \dots, c_n)$ とする.

3.4.2 ロバスト推定を用いた暫定カメラパラメータの推定

前節において算出された自然特徴点の候補位置 $\mathbf{c} = (c_0, c_1, \dots, c_n)$ に対し, 仮の追跡位置を決定し, LMeds 基準を用いたロバスト推定によって, 誤追跡を含まない少数の点を用いて暫定カメラパラメータを推定する. 以下に処理の流れを詳述する.

- (1) 第 $(f - 1)$ フレームにおいて存在する全ての自然特徴点に対して, 各特徴点 p の位置 $\mathbf{x}_{(f-1)p}$ の周辺をテンプレートとし, 第 f フレームにおいて, $\mathbf{x}_{(f-1)p}$ を中心とする一定幅の探索窓内で, マッチング誤差が最小となる点を仮の追跡位置とする.
- (2) 以下, ロバスト推定の i 回目の繰り返し処理を開始する. 追跡された n 個の特徴点からランダムに選択した k 個の特徴点 $P_i = \{p_{i1}, p_{i2}, \dots, p_{ik}\}$ により仮のカメラパラメータ \hat{M}_i を推定する.
- (3) 特徴点 p の三次元座標 S_p が \hat{M}_i を用いて画像上に投影された座標 $\hat{\mathbf{x}}_{fp}$ と画像上で観測される座標 \mathbf{x}_{fp} の距離の二乗誤差 (以下, 再投影誤差)

$$R_{ifp} = |\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2 \quad (8)$$

の中間値

$$RM_i = med(R_{if1}, R_{if2}, \dots, R_{ifn}) \quad (9)$$

を算出する.

(4) ステップ(2), (3), の処理を g 回繰り返す, LMedS 基準

$$LMedS = \min(RM_1, RM_2, \dots, RM_g) \quad (10)$$

を最小とするカメラパラメータを暫定カメラパラメータ \bar{M}_f とする.

上記の処理の中で, 特徴点の追跡結果 P_i を用いたカメラパラメータ \hat{M}_i の推定手法については3.5節で詳述する.

3.4.3 探索範囲を限定したマッチングによる自然特徴点の追跡

まず, 仮の追跡と同様に, 第 $(f-1)$ フレームにおいて存在する全ての自然特徴点に対して, 自然特徴点 p の画像上の位置 $x_{(f-1)p}$ の周辺をテンプレートとする. 次に, 第 f フレームにおいて, 自然特徴点 p の三次元座標 S_p が \bar{M}_f を用いて投影された座標 $\hat{x}_{(f-1)p}$ を中心とする一定幅の探索窓内で, マッチング誤差が最小となる点を自然特徴点 p の追跡位置とする. ただし, ここで用いる探索領域は, 仮の追跡に用いたものに比べて小さい範囲を指定することにより, 周辺類似パターンの存在による誤追跡を防ぐことが可能となる.

3.5 特徴点の追跡結果を用いたカメラパラメータの推定

本節では, 追跡結果を用いたカメラパラメータの推定手法について述べる. 本手法では, 式(8)で示した画像上の再投影誤差 R_{fp} を推定誤差として利用する. これにより, 第 f フレームにおけるカメラパラメータ M_f は, 次の誤差関数 E_f を最小化することで推定される.

$$E_f = \sum_p W_{fp} R_{fp} \quad (11)$$

ここで, W_{fp} は特徴点 p における重みであり, 3.7節で述べる特徴点 p の信頼度により決定する. また, カメラパラメータ M_f の自由度は6とし, カメラ座標系の基底ベクトルは単位直交条件を満たすものとする. このような制約の下での誤差関数 E_f の最小化は非線形最小化問題であり, 局所解や計算量の問題がある.

そこで、本研究では、まず式 (1) に示したカメラパラメータの自由度を 12 とし
て初期値を算出する。出口ら [7] によれば、カメラの外部パラメータを行列 $\mathbf{m} =$
 $(m_{11}, m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{31}, m_{32}, m_{33}, m_{34})^T$ で表現し、特徴点 p
の画像上での座標を $\mathbf{x}_p = (u_p, v_p)$ 、その世界座標系における同次座標値を $\mathbf{S}_p =$
 $(X_p, Y_p, Z_p, 1)$ としたとき、以下の関係が成り立つ。

$$\begin{pmatrix} \mathbf{S}_p & \mathbf{0} & -u_p \mathbf{S}_p \\ \mathbf{0} & \mathbf{S}_p & -v_p \mathbf{S}_p \end{pmatrix} \mathbf{m} = \mathbf{0} \quad (12)$$

これらを W_{fp} による重みを考慮して、 n 個の特徴点でまとめて表現すると、以下の
ようになる。

$$\begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} & -u_1 \mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{0} & -u_2 \mathbf{S}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{S}_n & \mathbf{0} & -u_n \mathbf{S}_n \\ \mathbf{0} & \mathbf{S}_1 & -v_1 \mathbf{S}_1 \\ \mathbf{0} & \mathbf{S}_2 & -v_2 \mathbf{S}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{S}_n & -v_n \mathbf{S}_n \end{pmatrix} \mathbf{m} = \mathbf{0} \quad (13)$$

$$\mathbf{W} = \begin{pmatrix} \sqrt{W_{f1}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \sqrt{W_{fn}} \end{pmatrix} \quad (14)$$

ここで、カメラパラメータのスケールを任意とし、 $m'_{ij} = m_{ij}/m_{34}$ 、 $m'_{34} = 1$ を用い
て式 (13) を変形すると、次式が得られる。

$$\mathbf{A} \mathbf{m}' = \mathbf{s} \quad (15)$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} & -u_1x_1 & -u_1y_1 & -u_1z_1 \\ \mathbf{S}_2 & \mathbf{0} & -u_2x_2 & -u_2y_2 & -u_2z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{S}_n & \mathbf{0} & -u_nx_n & -u_ny_n & -u_nz_n \\ \mathbf{0} & \mathbf{S}_1 & -v_1x_1 & -v_1y_1 & -v_1z_1 \\ \mathbf{0} & \mathbf{S}_2 & -v_2x_2 & -v_2y_2 & -v_2z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{S}_n & -v_nx_n & -v_ny_n & -v_nz_n \end{pmatrix} \quad (16)$$

$$\mathbf{m}' = \begin{pmatrix} m'_{11} \\ m'_{12} \\ m'_{13} \\ m'_{14} \\ m'_{21} \\ m'_{22} \\ m'_{23} \\ m'_{24} \\ m'_{31} \\ m'_{32} \\ m'_{33} \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} \sqrt{W_{f1}}u_1 \\ \sqrt{W_{f2}}u_2 \\ \vdots \\ \sqrt{W_{fn}}u_n \\ \sqrt{W_{f1}}v_1 \\ \sqrt{W_{f2}}v_2 \\ \vdots \\ \sqrt{W_{fn}}v_n \end{pmatrix} \quad (17)$$

これに対して擬似逆行列を用いれば, $\mathbf{m}' = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{s}$ となる. ここで, m_{34} が未確定であるため, カメラパラメータ M のスケールに自由度が残るが, スケールに関係無く特徴点の画像上への投影座標は一定である. よって, ここでは m_{34} を任意に決定することで \mathbf{m} を求める.

しかし, このままではカメラパラメータにスケールの自由度が残り, また自由度が12であるためにカメラ姿勢のパラメータである θ, δ, γ を決定できない. そこで, 観測誤差が画像上の誤差であることに着目すると, 推定されたカメラ位置の誤差が主としてカメラの光軸方向に含まれることが言えるため, 真のカメラ位置が推定された光軸上に存在すると仮定し, 線形演算によってカメラパラメータの

自由度を6に補正する.

次に, 補正されたカメラパラメータを初期値として勾配法により E_f を最小化する. この非線形最小化問題の初期値は先に行った線形演算による真値に近い推定値であるため, 局所解に陥ることを回避し, 少ない演算回数で大域最小解を求めることができると考えられる. これにより, 追跡結果から E_f を最小化する M_f を求めることができる.

3.6 自然特徴点の三次元位置の推定

提案手法では, 自然特徴点の三次元位置を毎フレームで推定し, 更新する. 推定された自然特徴点の三次元位置は, 次フレームにおけるカメラパラメータ推定に利用される. 第 $(f-k)$ フレームから第 f フレームまで追跡されている自然特徴点 p の三次元位置 $S_p = (X_p, Y_p, Z_p)$ は, その画像上の座標系列 $\mathbf{x}_{(f-k)p}, \mathbf{x}_{(f-k+1)p}, \dots, \mathbf{x}_{fp}$ と, カメラパラメータ系列 $M_{(f-k)}, M_{(f-k+1)}, \dots, M_f$ から推定される. ここでは, 図6に示すように, 自然特徴点が追跡された全フレームにおいて, 自然特徴点の画像上の座標とカメラの投影中心を結ぶ直線を考える. これらの直線は追跡誤差や量子化誤差の影響により一点で交わらないため, 提案手法では, これらの直線との距離の二乗和が最小となる点を推定位置とする. この点は, 以下の手順によって算出される.

第 f フレームにおけるカメラの投影中心と自然特徴点 p の世界座標における三次元座標を結ぶベクトルの単位ベクトルを $\mathbf{vz}_f = (x_f, y_f, z_f)$ とする. まず \mathbf{vz}_f に対して直交する単位ベクトル \mathbf{vy}_f を以下の式により決定する.

$$\mathbf{vy}_f = \left(\frac{z_f}{\sqrt{x_f^2 + z_f^2}}, 0, -\frac{x_f}{\sqrt{x_f^2 + z_f^2}} \right) \quad (18)$$

また \mathbf{vy}_f と \mathbf{vz}_f のどちらにも直交する単位ベクトルを $\mathbf{vx}_f = \mathbf{vy}_f \times \mathbf{vz}_f$ とする. このとき, 自然特徴点 p と第 n フレームまでの各フレームの投影中心と自然特徴点 p を結ぶ直線の距離の二乗和 R は, $\mathbf{vx}_f, \mathbf{vy}_f$ および, 世界座標におけるカメラ位置 $\mathbf{t}_f = (tx_f, ty_f, tz_f)$ を用いて以下の式により算出できる.

$$R = \sum_{i=f-k}^f (rx_i^2 + ry_i^2) \quad (19)$$

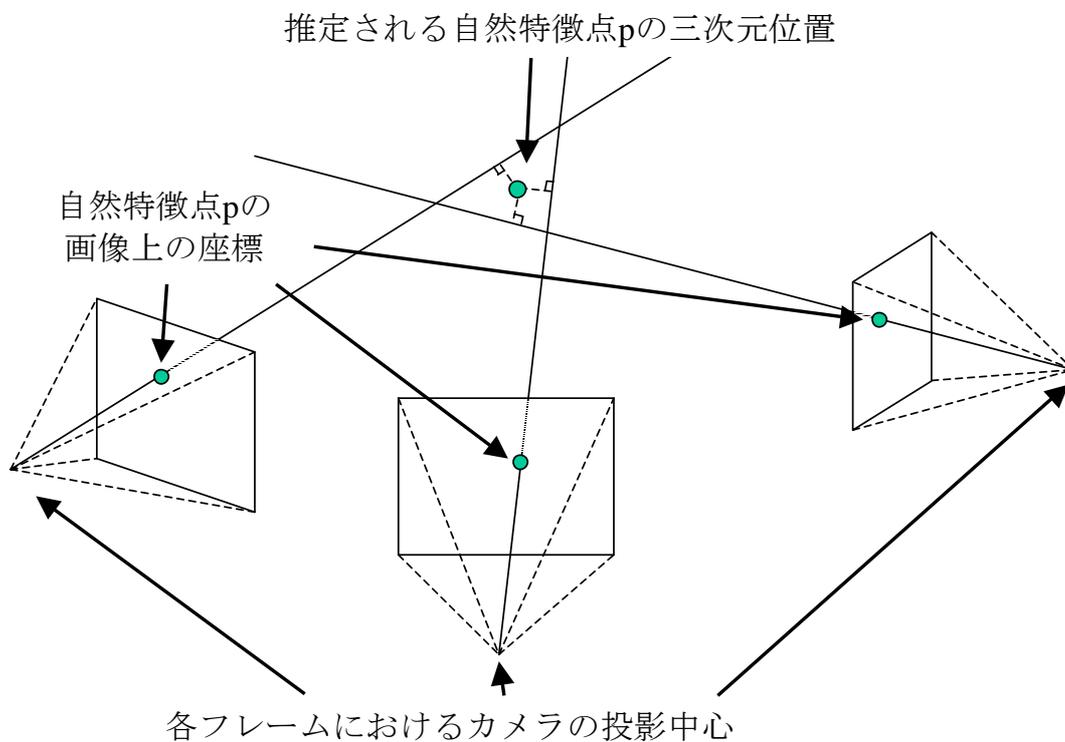


図 6 推定される自然特徴点の三次元位置

$$\mathbf{r} = \begin{pmatrix} rx_{(f-k)} \\ rx_{(f-k+1)} \\ \vdots \\ rx_f \\ ry_{(f-k)} \\ ry_{(f-k+1)} \\ \vdots \\ ry_f \end{pmatrix} = \begin{pmatrix} \mathbf{vX}_{(f-k)} & tx_{(f-k)} \\ \mathbf{vX}_{(f-k+1)} & tx_{(f-k+1)} \\ \vdots & \vdots \\ \mathbf{vX}_f & tx_f \\ \mathbf{vY}_{(f-k)} & ty_{(f-k)} \\ \mathbf{vY}_{(f-k+1)} & ty_{(f-k+1)} \\ \vdots & \vdots \\ \mathbf{vY}_f & ty_f \end{pmatrix} \begin{pmatrix} X_p \\ Y_p \\ Z_p \\ 1 \end{pmatrix} \quad (20)$$

これを変形すると、次式が得られる。

$$\mathbf{r} + \mathbf{t} = \mathbf{CS}_p \quad (21)$$

$$\mathbf{t} = \begin{pmatrix} -tx_{(f-k)} \\ -tx_{(f-k+1)} \\ \vdots \\ -tx_f \\ -ty_{(f-k)} \\ -ty_{(f-k+1)} \\ \vdots \\ -ty_f \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{vX}_{(f-k)} \\ \mathbf{vX}_{(f-k+1)} \\ \vdots \\ \mathbf{vX}_f \\ \mathbf{vY}_{(f-k)} \\ \mathbf{vY}_{(f-k+1)} \\ \vdots \\ \mathbf{vY}_f \end{pmatrix} \quad (22)$$

ここで、擬似逆行列を用いれば、 R を最小とする三次元座標値 $\mathbf{S}_p = (X_p, Y_p, Z_p)$ は、 $\mathbf{S}_p = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{t}$ により得られる。

3.7 特徴点の信頼度の算出

本節では、次フレーム(第 $(f+1)$ フレーム)においてカメラパラメータ推定の重みとして用いるための特徴点の信頼度の算出方法について述べる。第 $(f+1)$ フレームでの処理において、追跡される特徴点 p の画像上の座標 $\mathbf{x}_{(f+1)p}$ は、追跡誤差により p の三次元位置 \mathbf{S}_p を画像上に投影した座標 $\hat{\mathbf{x}}_{(f+1)p}$ に一致しない。そこで、この追跡誤差が $\hat{\mathbf{x}}_{(f+1)p}$ 周辺に分散 σ_p^2 をもつガウス分布を成すものとしてモデル化を行う。

このとき、 $\mathbf{x}_{(f+1)p}$ が真値である確率 $p(\mathbf{x}_{(f+1)p})$ は次式のようになる。

$$p(\mathbf{x}_{(f+1)p}) = \frac{1}{2\pi\sigma_p^2} \exp\left(-\frac{|\mathbf{x}_{(f+1)p} - \hat{\mathbf{x}}_{(f+1)p}|^2}{2\sigma_p^2}\right) \quad (23)$$

これを用い、第 $(f+1)$ フレームにおけるすべての特徴点の確率密度関数の積を以下に示す。

$$P_{(f+1)} = \prod_p p(\mathbf{x}_{(f+1)p}) \quad (24)$$

このような確率密度関数の積 $P_{(f+1)}$ を最大とするカメラパラメータ $\mathbf{M}_{(f+1)}$ は、

$$EM_{(f+1)} = \sum_p \log(2\pi\sigma_p^2) + \sum_p \frac{|\mathbf{x}_{(f+1)p} - \hat{\mathbf{x}}_{(f+1)p}|^2}{2\sigma_p^2} \quad (25)$$

を最小化することにより得られる。ただし、式(25)の右辺の第一項は定数であるため実際には以下の関数を最小化すればよい。

$$E\hat{M}_{(f+1)} = \sum_p \frac{|\mathbf{x}_{(f+1)p} - \hat{\mathbf{x}}_{(f+1)p}|^2}{2\sigma_p^2} \quad (26)$$

本手法では、 σ_p^2 として第 f フレームまでの再投影誤差の分散を用い、式(11)と式(26)を比較することにより、第 $(f-k)$ フレームから第 f フレームまで追跡されている特徴点 p の追跡の信頼度 $W_{(f+1)p}$ を次式のように定義する。

$$W_{(f+1)p} = \frac{1}{2\sigma_p^2} = \frac{k+1}{2} \left\{ \sum_{i=f-k}^f |\mathbf{x}_{ip} - \hat{\mathbf{x}}_{ip}|^2 \right\}^{-1} \quad (27)$$

3.8 自然特徴点の追加と削除

自然特徴点の追加は、3.4.1節によって得られる自然特徴点の候補位置に対して、画像上の情報のみを用いて追跡を行い、一定フレーム以上追跡された追跡結果のなかから、以下の条件を全て満たすものを選択することで実現する。

- 信頼度が一定値以上
- 各画像間のマッチング誤差が一定値以下
- Harris オペレータにより得られる特徴量が一定値以上
- 推定される三次元位置と各フレームのカメラの投影中心を結ぶ直線の成す最大角が一定値以上

また、以下のどちらかの条件を満たす特徴点を削除する。

- 信頼度が一定値以下
- 各画像間のマッチング誤差が一定値以上

3.9 動画像系列全体での最適化

これまでの処理は逐次的に行うため各フレームにおける処理は短時間で行えるが、誤差が蓄積するという問題がある。そこで、カメラパラメータ M_f と自然特徴点 p の三次元位置 S_p をパラメータとし、最後に全フレームに対して最適化を行う。最適化の誤差関数には次式を用い、 M_f と S_p の初期値には逐次処理によって得られた推定結果を用いる。

$$E = \sum_f \sum_p W_p |\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2 \quad (28)$$

ここで、 $\hat{\mathbf{x}}_{fp}$ は式 (5) によって特徴点の三次元座標 S_p がカメラパラメータ M_f の画像上に投影された座標であり、 W_p には式 (27) に示す自然特徴点 p の信頼度を用いる。ただし、追跡中に削除された自然特徴点に関しては、削除されたフレームから一定フレーム前までの追跡の信頼度が低いといえるため、その追跡結果は利用しない。この最適化は非線形最小化問題となるが、初期値が真値に近ければ解は大域最小解に陥りやすくなり、また高速に求めることができると考えられる。

4. 実験

提案手法が、現実環境を撮影した自然特徴点の出現や消失のおこる多数の画像系列から、精度の良い三次元復元を行えることを示すために、表 2 に示す実験条件において、(1) 形状が比較的単純で撮影時間が短い屋内環境、(2) 複雑な形状を持ち撮影時間が長い屋外環境、を対象として実験を行った。本実験では、カメラの内部パラメータは Tsai の手法 [32] によりあらかじめ得た値を利用し、提案手法により表 3 に示す閾値を用いてカメラの外部パラメータと自然特徴点の三次元位置の推定を行った。

表 2 実験環境

	(1) 屋内環境における実験	(2) 屋外環境における実験
カメラ	Sony DCR-VX 1000	Sony DSR-DP150
ワイドレンズ	無し	Sony VCL-HG0758(0.7倍)
画像の解像度	720×240画素	720×480画素
撮影時間	8.9秒	40秒
撮影枚数	267枚	599枚
基準マーカの追跡	手動	色領域による自動追跡
計算機	PC (CPU : Pentium III 1GHz, Memory : 512MB)	

4.1 屋内環境における実験

図 7 に示す静的な実環境を手持ちの CCD カメラ (Sony DCR-VX 1000) で 8.9 秒間撮影した動画画像 267 枚 (720×240 画素) を入力とした。提案手法では基準マーカは色や形状の情報を用いて追跡することを想定しているが、本実験では各フレームにおける基準マーカの画像上の位置は既知とした。

図 8 に、実験で指定した基準マーカを印で、提案手法を用いて追跡された自然特徴点を×印で示す。これより、画像中のエッジのコーナ部分で自然特徴点が多

表 3 実験に用いたパラメータ

自然特徴点追跡用

	(1) 屋内実験	(2) 屋外実験
パターンテンプレートのサイズ (画素)	9×9	12×12
仮追跡の探索範囲 (画素)	30×30	50×50
限定された探索範囲 (画素)	18×18	33×33

自然特徴点の追加の評価尺度

	(1) 屋内実験	(2) 屋外実験
信頼度の最小値 (分散の逆数)	$\frac{1}{6}$	$\frac{1}{4}$
信頼度の最小値 (分散の逆数)	$\frac{1}{6}$	$\frac{1}{4}$
投影中心と点を結ぶ直線間の成す最小角 (度)	5.0	4.0
最低追跡フレーム数 (枚)	12	30
フレーム間のテンプレートの最低相関値	0.99	0.99

自然特徴点の削除の評価尺度

	(1) 屋内実験	(2) 屋外実験
信頼度の最小値 (分散の逆数)	$\frac{1}{8}$	$\frac{1}{4}$
フレーム間のテンプレートの最低相関値	0.99	0.99

く検出・追跡されており、位置ずれの少ない追跡が実現できていることが確認できる。

図 9に、推定されたカメラパラメータを用いて、CGで描かれた箱を入力画像に重畳表示した結果 (Match Move) を示す。同図に示すように、重畳表示されたCGは現実環境の一定の位置からずれることなく表示されていることから、カメラパラメータの推定に大きな誤差が含まれていないと言える。

次に、カメラの位置・姿勢と、自然特徴点の三次元座標および撮影対象の形状の推定結果を図 10に示す。図中の実線は、推定されたカメラの移動軌跡であり、四角錐は15フレーム毎のカメラの姿勢を示す。復元形状は、信頼度が一定の閾値を超える自然特徴点の三次元座標をカメラ座標系に投影した画像上の座標に対し、Delaunayの三角分割法[37]を適用して三角パッチを自動生成し、数個のカメラ画像から得られるパッチの統合処理を行うことにより得た。

図 10から分かるように、カメラ位置の軌跡は滑らかに推定されており、また基準マーカが観測できないカメラ位置においても、カメラの位置・姿勢を復元することができた。しかし、追跡されている自然特徴点の数が少ない初期の数十フレームの間ではカメラ位置の推定結果に誤差が含まれていた。また、復元形状に関しては面を構成するために必要な全ての自然特徴点が復元されていないため、本来の形状と異なる面が構成されているところも見られたが、カメラ位置と異なる視点においても違和感の無い画像が生成できた。

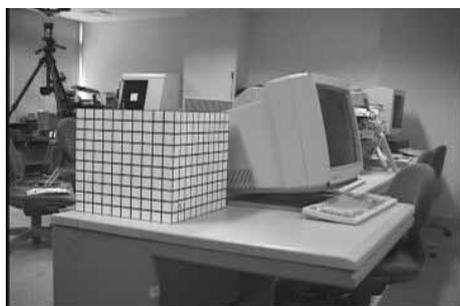
実験にはPC (CPU : Pentium III 1GHz, Memory : 512MB) を用い、図 7の8.9秒の入力動画像に対し、図 2(A)の逐次的な処理により全267フレームで自然特徴点の三次元位置とカメラの位置・姿勢を推定するのに要した計算時間は98秒であった。次に図 2(B)の画像系列全体での最適化処理における、処理の繰り返し回数と式 (28)の誤差関数 E の関係を図 11に示す。横軸は、最適化処理の繰り返し回数、縦軸は、逐次処理によって得られる式 (28)の E の初期値に対するパーセンテージを表す。同図に示すように、500回の繰り返し処理により最適化を行ったところ、 E は初期値の59.9%となり、この処理に130秒を要した。同様の条件で5000回の繰り返し処理を行ったところ初期値の58.0%となり、約1200秒を要した。



第1フレーム



第160フレーム



第40フレーム



第200フレーム



第80フレーム



第240フレーム



第120フレーム



第266フレーム

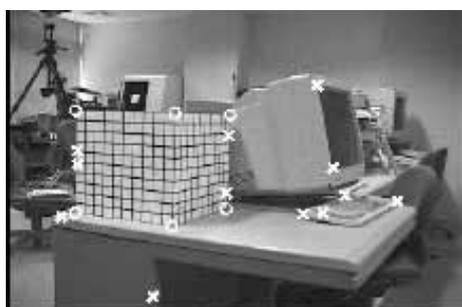
図7 屋内環境での実験における入力画像



第1フレーム



第160フレーム



第40フレーム



第200フレーム



第80フレーム



第240フレーム

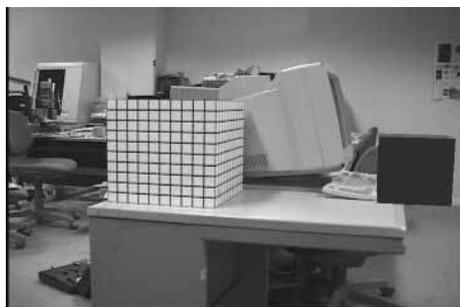


第120フレーム



第266フレーム

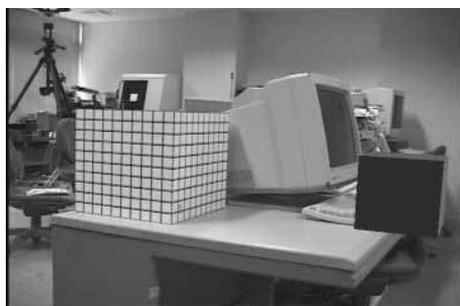
図8 屋内環境での実験における特徴点の追跡結果



第1フレーム



第160フレーム



第40フレーム



第200フレーム



第80フレーム



第240フレーム



第120フレーム



第266フレーム

図9 屋内環境での実験における MatchMove

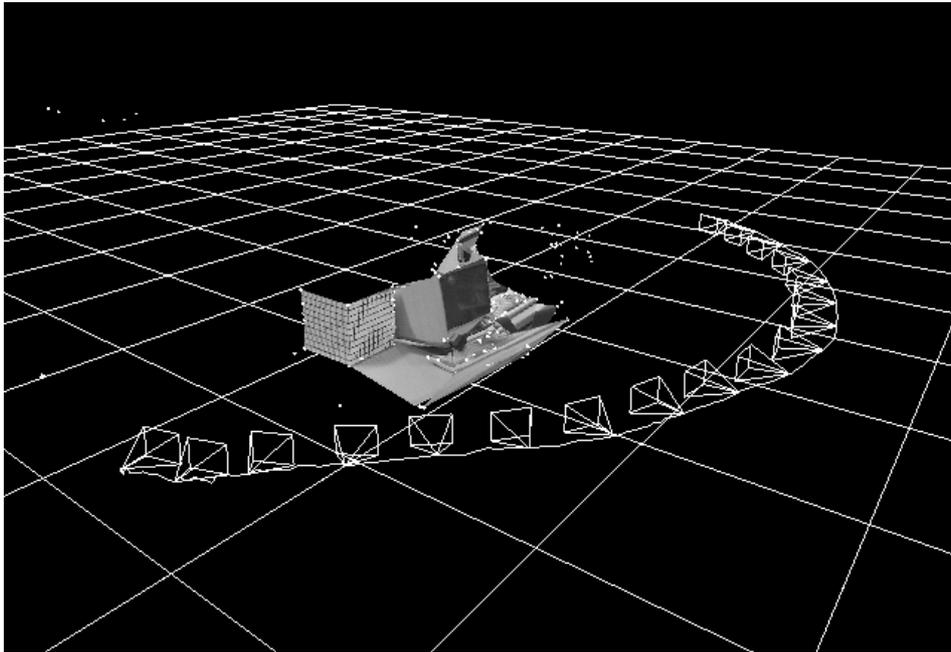
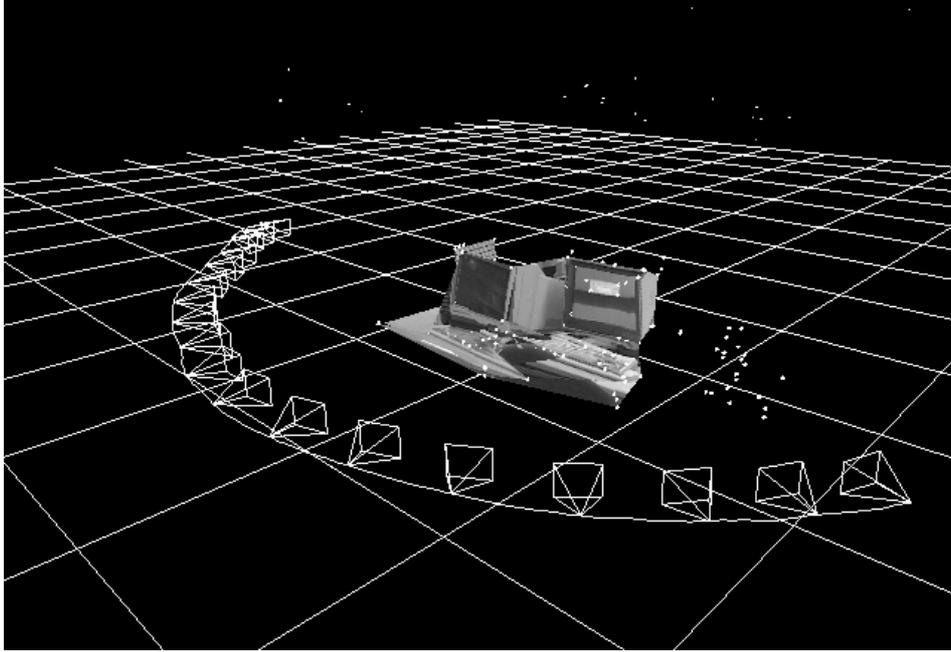


図 10 カメラの位置・姿勢と自然特徴点の三次元位置および面の構成結果 (屋内環境の場合)

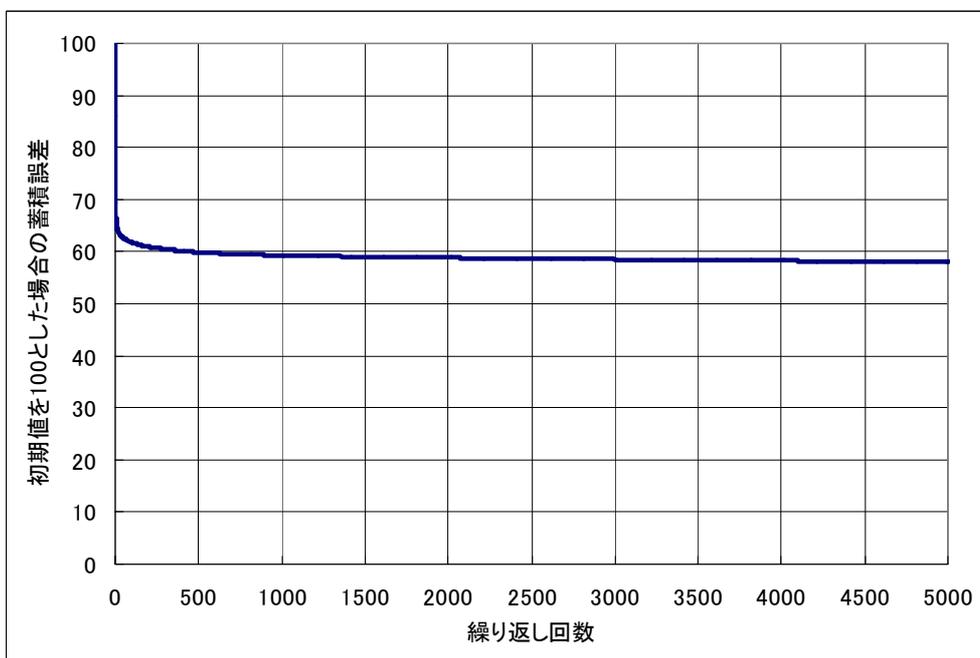


図 11 動画像全体での最適化による累積誤差の推移 (屋内環境の場合)

4.2 屋外環境における実験

広角レンズ (Sony VCL-HG0758) を取り付けた手持ちの CCD カメラ (Sony DSR-DP-150) で図 12 に示す静的な屋外環境を 40 秒間撮影し、得られた画像 599 枚 (720×480 画素, プログレッシブ撮影) を入力として実験を行った。

実験では、図 13 の第 1 フレームにおいて 印で示す 6 点を基準マーカとして指定し、第 2 フレーム以降は色領域による自動追跡を行った。同図中に、基準マーカと自然特徴点 (× 印) の自動追跡の結果を示す。屋内環境での実験と同様に、画像中のエッジのコーナ部分で自然特徴点が多く検出・追跡された。

図 14 に、推定されたカメラパラメータを用いて、CG で描かれた箱を入力画像に重畳表示した結果 (Match Move) を示す。このような複雑な撮影対象を長時間撮影した入力動画像に対しても、屋内実験と同様、重畳表示された CG は現実環境の一定の位置からずれることなく表示されており、カメラパラメータの推定には大きな誤差が含まれていないと言える。

推定されたカメラの位置・姿勢と、自然特徴点の三次元座標および撮影対象の形状を図 15 に示す。形状の復元には屋内環境における実験と同様の手法を用いたが、撮影対象の屋根の部分に自然特徴点を検出されなかったため、本来と異なる面が構成され、カメラ位置と異なる視点から生成される画像に大きな歪みが生じた。

計算機は屋内環境における実験と同じものを使用し、図 12 の 40 秒 (全 599 フレーム) の入力動画像に対して、図 2(A) の逐次的な処理に要した計算時間は 514 秒であった。また、図 16 に示すように、図 2(B) の画像系列全体での最適化処理において 500 回の繰り返し処理によって式 (28) の誤差関数 E は初期値の 81.5% となり、約 580 秒を要した。同様に 5000 回の繰り返し処理においては誤差関数 E は初期値の 80.0% となり約 4700 秒を要した。

4.3 考察

提案手法では、Harris オペレータにより、追跡の容易な特徴点を画像上で検出し、ロバスト推定の手法により、自然特徴点の三次元位置を考慮した追跡を行う。屋内環境および屋外環境における実験より、画像上のエッジのコーナなどの追跡の容易な点で自然特徴点が多く検出され、周辺に類似パターンが存在する領域に



第1フレーム



第340フレーム



第85フレーム



第425フレーム



第170フレーム



第510フレーム

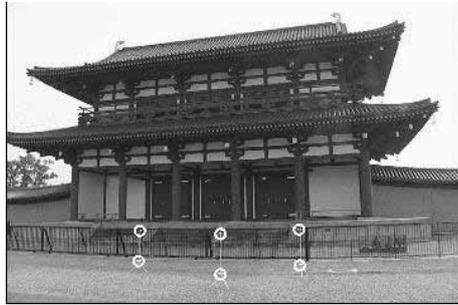


第255フレーム



第599フレーム

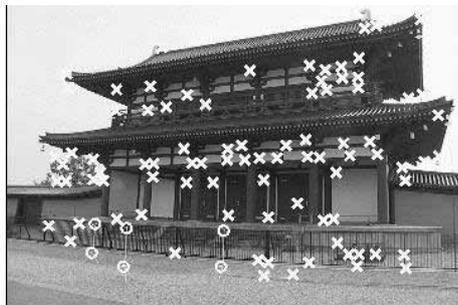
図 12 屋外環境での実験における入力画像



第1フレーム



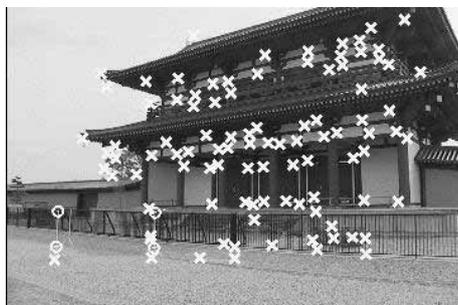
第340フレーム



第85フレーム



第425フレーム



第170フレーム



第510フレーム



第255フレーム



第599フレーム

図 13 屋外環境での実験における自然特徴点の追跡結果



第1フレーム



第340フレーム



第85フレーム



第425フレーム



第170フレーム



第510フレーム



第255フレーム



第599フレーム

図 14 屋外環境での実験における MatchMove

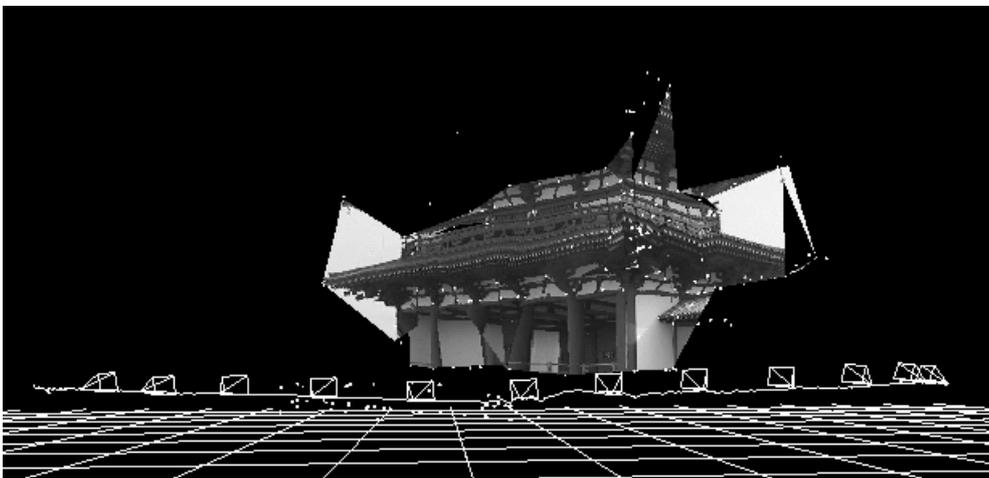
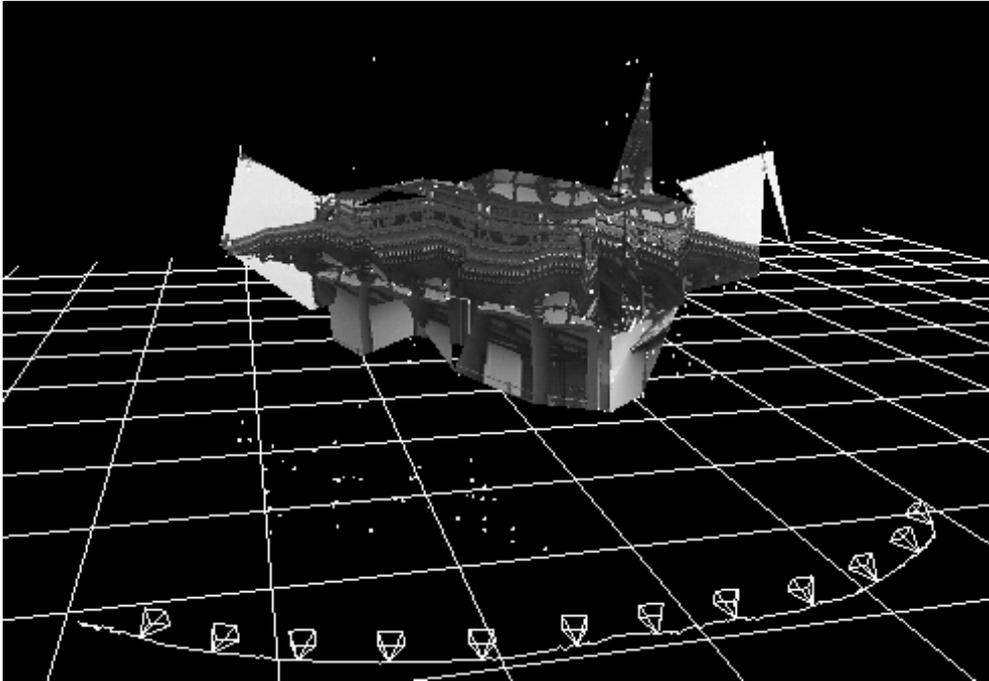


図 15 カメラの位置・姿勢と自然特徴点の三次元位置および面の構成結果 (屋外環境の場合)

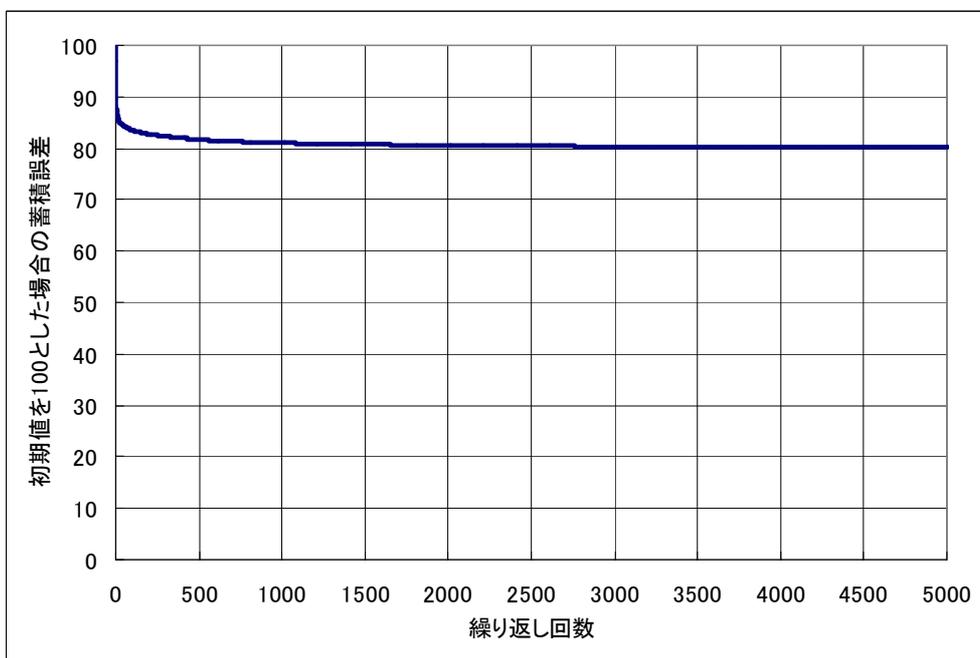


図 16 動画像全体での最適化による累積誤差の推移 (屋外環境の場合)

においても、位置ずれや誤対応の少ない、精度の高い自動追跡が行われていることを確認した。また、追跡に失敗し、信頼度が低下した自然特徴点は自動的に削除されており、自然特徴点の出現と消失に対応できていることを確認した。さらに、MatchMoveにより、基準マーカが観測できない視点位置においても大きな誤差を伴うことなくカメラパラメータが推定されていることが確認できた。

復元対象の形状に関しては、例えば屋外実験における建物の屋根の部分のように、周辺類似パターンが存在する領域では、自然特徴点が追跡されず、画像上での三角分割の結果を用いた面の構成においてはこのような領域で奥行きが正しく復元されず、歪みがみられた。このため、更に精度の良い形状の復元を行うためには、より密な対象の復元を行うことが必要であると考えられる。

図 11,16から、動画像系列全体を用いた最適化によって、蓄積誤差が減少しており、また、少ない繰り返し回数によって、最適解に近い値を推定できることが確認できる。両図において屋内実験と屋外実験を比較すると、屋内実験は最適化後の累積誤差が小さくなっている。これは、屋内環境では、オクルージョンによって追跡中に自然特徴点が消失した場合でも、信頼度が閾値以下に低下するまでの数フレーム間に渡って自然特徴点が追跡されるため、逐次推定における累積誤差が増加したものと考えられる。動画像系列全体での最適化処理においては、これら信頼度が低いと考えられる一定フレーム間の追跡結果を使用しない。このため、比較的特徴点の消失が多くみられた屋内実験において誤差の評価値の減少が顕著になった。

5. まとめ

本論文では、三次元位置関係および色と形状が既知の基準マーカを複数利用し、初期フレームにおいてこれらを画像上で指定し、自然特徴点と共に自動追跡を行うことで、動画像からシーンの三次元復元を行う手法を提案した。このような事前知識は、従来手法が必要とする事前知識に比べて容易に得ることができ、また、復元対象およびカメラの動きに関する制約は小さいといえる。提案手法では、毎フレーム、自然特徴点の三次元位置とカメラパラメータを復元することで、動画像を用いた従来手法において問題となる特徴点の出現・消失に対応し、カメラモデルとして透視投影モデルを用いることにより、奥行きの大い対象も復元することができる。

提案手法を用いて三次元復元実験を行った。実験では屋内環境および屋外環境を撮影した動画像を入力として、カメラパラメータと自然特徴点の三次元位置を推定し、推定されたカメラパラメータに大きな誤差が含まれていないことを MatchMove によって示した。また、動画像系列全体での最適化処理を行うことにより、蓄積誤差が削減され、より精度の高い復元が行われたことを示した。これにより、上述のような提案手法の有効性を確認することができた。しかし、実験で用いた Delaunay の三角分割法を利用した面の構成では復元された形状に歪みが見られた。

今後は、本手法で推定されたカメラパラメータにより算出される各画像間のエピポラ拘束を利用し、密な奥行き推定を行うことで、実験において自然特徴点が追跡されなかった領域においても奥行きを算出し、より精度の高い形状を復元できると考えられる。また、より長時間、自然特徴点を追跡することができれば、更なる復元精度の向上が期待できるため、全方位カメラを用いた入力画像に対する本手法の適用を検討する予定である。これにより、さらに広域な環境を復元することが可能になるとと思われる。

謝辞

本研究の全過程を通して、懇切なる御指導、御鞭撻を賜ったソフトウェア基礎講座 横矢 直和教授に心より深謝致します。

本研究の遂行にあたり、御助言頂いた像情報処理学講座 千原 國宏教授に厚く御礼申し上げます。

本研究を進めるにあたり、終始、暖かい御指導をして頂いたソフトウェア基礎講座 竹村 治雄助教授に深く感謝いたします。

そして、本研究を通じて、有益な御助言を頂いたソフトウェア基礎講座 山澤 一誠助手に厚く感謝いたします。

本研究の遂行において多大なる御助言、御鞭撻を賜り、終始御指導を頂いた 神原 誠之氏に心から感謝いたします。また、日頃から有益な御意見を頂き御指導して頂いた、町田 貴史氏、ならびに、本研究において、多大な協力を頂いた寺田 智裕氏、富手 要氏、ソフトウェア基礎講座の諸氏に深く御礼申し上げます。

最後に、日頃より暖かく支えて下さったソフトウェア基礎講座 事務補佐員 福永 博美女史、ならびに、北川 知代女史に厚く感謝いたします。

参考文献

- [1] N. Yokoya, T. Shakunaga and M. Kanbara: “Passive Range Sensing Techniques: Depth from Images,” *IEICE Trans. Inf. and Syst.*, Vol. E82-D, No. 3, pp. 523–533, 1999.
- [2] C. Tomasi and T. Kanade: “Shape and Motion from Image Streams under Orthography: A Factorization Method,” *Int. Journal of Computer Vision*, Vol. 9, No. 2, pp. 137–154, 1992.
- [3] H. S. Sawhney, Y. Guo, J. Asmuth and R. Kumar: “Multi-view 3D Estimation and Application to Match Move,” *Proc. IEEE Workshop on Multi-view Modeling and Analysis of Visual Scenes*, pp. 21–28, 1999.
- [4] 上原, 塩崎, 全: “数値地図と動画像解析による都市空間モデリング”, *画像の認識・理解シンポジウム (MIRU 2000) 講演論文集*, Vol. II, pp. 373–378, 2000.
- [5] M. Pollefeys, R. Koch, M. Vergauwen, A. A. Deknuydt and L. J. V. Gool: “Three-dimensional Scene Reconstruction from Images,” *Proc. SPIE*, Vol. 3958, pp. 215–226, 2000.
- [6] R. Szeliski and S. B. Kang: “Recovering 3D Shape and Motion from Image Streams Using Non-linear Least Squares,” *Journal of Visual Communication and Image Representation*, Vol. 6, No. 1, pp. 10–28, 1994.
- [7] 出口: “射影幾何学による PnP カメラ補正問題の統一的解法”, *情報シンポジウム*, Vol. 90, pp. 41–50, 1990.
- [8] B. K. P. Horn: “Obtaining Shape from Shading Information,” in *The Psychology of Computer Vision* (Ed. by P. H. Winston), New York, McGraw-Hill, pp. 115–155, 1975.
- [9] R. J. Woodham: “Photometric Method for Determining Surface Orientation from Multiple Images,” *Opt. Eng.*, Vol. 19, No. 1, pp. 139–144, 1981.

- [10] K. Ikeuchi: “Determining Surface Orientation of Specular Surfaces by using The Photometric Stereo Method,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. PAMI-3, No. 6, pp. 661–669, 1981.
- [11] S. K. Nayar: “Shape from Focus System,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 302–308, 1992.
- [12] H. N. Nair and C. V. Stewart: “Robust Focus Ranging,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 309–314, 1992.
- [13] S. K. Nayar and Y. Nakagawa: “Shape from Focus,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 8, pp. 824–831, 1994.
- [14] S. Hiura and T. Matsuyama: “Depth Measurement by The Multi-focus Camera,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 953–959, 1998.
- [15] A. N. Rajagopalan and S. Chaudhuri: “Optimal Selection of Camera Parameters for Recovery of Depth from Defocused Images,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 219–224, 1997.
- [16] J. Y. Zheng: “Acquiring 3-D Models from Sequences of Contours,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 163–178, 1994.
- [17] W. Niem and J. Wingbermühle: “Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera,” *Proc. Int. Conf. on Recent Advances in 3D Imaging and Modeling*, pp. 173–180, 1997.
- [18] 岡谷, 出口: “視点位置が不明な画像系列における物体の輪郭形状の重ね合わせ”, *画像の認識・理解シンポジウム (MIRU 2000) 講演論文集*, Vol. II, pp. 49–54, 2000.

- [19] Y. Ohta and T. Kanade: “Stereo by Intra- and Inter- Scanline Search Using Dynamic Programming,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No. 2, pp. 139–154, 1985.
- [20] 横矢: “多重スケールでの正則化によるステレオ画像からの不連続性を保存した曲面再構成”, 信学論, Vol. J76-D-II, No. 8, pp. 1667–1675, 1993.
- [21] N. Ayache and F. Lustman: “Fast and Reliable Passive Trinocular Stereo Vision,” Proc. 1st Int. Conf. on Computer Vision, pp. 422–426, 1987.
- [22] V. J. Milenkovic and T. Kanade: “Trinocular Vision Using Photometric and Edge Orientation constraints,” Proc. Image Understanding Workshop, pp. 163–175, 1985.
- [23] J. Poleman and T. Kanade: “A Paraperspective Factorization Method for Shape and Motion Recovery,” Tech. Rep. CMU-CS-93-219, Carnegie-Mellon Univ., 1993.
- [24] 戎野, 山澤, 竹村, 横矢: “全方位ステレオ画像センサによる実環境の仮想環境への取り込み”, 信学技報, MVE99-82, pp. 37–44, 1999.
- [25] T. Mukai and N. Ohnishi: “The Recovery of Object Shape and Camera Motion Using a Sensing System with a Video Camera and Gyro Sensor,” Proc. 7th Int. Conf. on Computer Vision, pp. 411–417, 1999.
- [26] 出口: “透視投影条件下での因子分解法による多視点画像からの形状復元”, 情処研報 CVIM106-6, 1997.
- [27] 植芝, 富田: “初期モデルとしてアフィン投影を用いた多視点透視投影画像からの運動と形状の復元”, 情処研報 CVIM107-1, 1997.
- [28] 浮田, 尺長: “透視投影因子分解法による広範囲画像系列からの3次元物体モデル生成”, 信学技報 PRMU97-276, 1998.

- [29] L. Quan and T. Kanade: “A Factorization Method for Affine Structure from Line Correspondence,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 803–808, 1996.
- [30] D. D. Morris and T. Kanade: “A Unified Factorization Algorithm for Points, Lines Segments and Planes with Uncertainty Models,” 画像の認識・理解シンポジウム (MIRU '98) 講演論文集, Vol. I, pp. 401–406, 1998.
- [31] 右田, 天野, 浅田: “非線形最適化計算を用いた運動からの形状復元”, 画像の認識・理解シンポジウム (MIRU 2000) 講演論文集, Vol. II, pp. 199–204, 2000.
- [32] 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.
- [33] Tomasi and Kanade: “Detection and Tracking of Point Features,” Tech. Rep. CMU-CS-91-132, Carnegie-Mellon Univ., 1991.
- [34] 藤井, 神原, 竹村, 横矢: “拡張現実のためのジャイロセンサを併用したステレオカメラによる位置合わせ”, 信学技報, MVE99-59, 2000.
- [35] C. Harris and M. Stephens: “A Combined Corner and Edge Detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.
- [36] C. Schmid, R. Mohr and C. Bauckhage: “Comparing and Evaluating Interest Points,” Proc. 6th Int. Conf. on Computer Vision, pp. 230–235, 1998.
- [37] P. Heckbert Ed.: Graphics Gems IV, pp. 47–59, Academic Press, 1994.