

NAIST-IS-DT0161018

博士論文

複数の動画像を用いたカメラパラメータ推定に基づく  
屋外環境の三次元モデル化に関する研究

佐藤 智和

2003年 3月 24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  
博士(工学) 授与の要件として提出した博士論文である。

佐藤 智和

審査委員： 横矢 直和 教授  
千原 國宏 教授  
山澤 一誠 助教授

# 複数の動画像を用いたカメラパラメータ推定に基づく 屋外環境の三次元モデル化に関する研究\*

佐藤 智和

## 内容梗概

動画像からの屋外環境の三次元モデル化は、物体認識、景観シミュレーション、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、現在このような分野で用いる三次元モデルは主として手動で作成されるため、作成コストが高く自動化が求められている。これらを自動化するための試みとして、従来から動画像を用いる三次元復元手法に関する研究が盛んであるが、復元範囲や復元精度に問題があるために、広域で複雑な屋外環境を精度良く復元するには至っていない。本論文では、複数の動画像と三次元位置が既知の基準マーカを入力として用いることで、特徴点追跡に基づくカメラパラメータの自動推定と多視点ステレオ視により、広域で複雑な屋外環境を精度良く復元する手法を提案する。1章では、現実環境の三次元モデル化に関する従来研究を概観し、本研究の位置づけと研究方針を明確にする。2章では、三次元位置が既知の基準マーカと自然特徴点を画像上で追跡し、カメラの移動パラメータを精度良く推定する手法について述べる。また、現実環境を撮影した動画像を入力とし、評価実験を行うことで提案手法の有効性を確認する。3章では、カメラパラメータが推定された複数の動画像系列を用いて、拡張マルチベースラインステレオ法により、各フレームでの奥行き情報を密に推定し、それらをボクセル空間に統合することで、三次元モデルを復元する手法を提案する。また、実際に複数の動画像系列を用いた屋外環境の三次元

---

\* 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 博士論文, NAIST-IS-DT0161018, 2003年3月24日.

モデル化結果を示す．最後に，4章で本研究を総括するとともに今後の展望について述べる．

キーワード

動画像解析，カメラパラメータ推定，マルチベースラインステレオ，三次元復元



# Reconstruction of 3-D Models of Outdoor Scenes Based on Estimating Extrinsic Camera Parameters from Multiple Image Sequences\*

Tomokazu Sato

## Abstract

Three-dimensional (3-D) models of outdoor scenes can be widely used in a number of fields such as object recognition, navigation, scenic simulation, and mixed reality. Such 3-D models are often made manually with high costs, so that automatic 3-D reconstruction has been widely investigated. In related works, a dense 3-D model is generated by using a stereo method. However, such approaches cannot use several hundred images together for dense depth estimation because it is difficult to accurately calibrate a large number of cameras. This thesis proposes a novel dense 3-D reconstruction method that first estimates extrinsic camera parameters of a hand-held video camera, and then reconstructs a dense 3-D model of a scene using an extended multi-baseline stereo technique. Chapter 1 gives a perspective of the study in the area of 3-D reconstruction. Chapter 2 describes a method for estimating extrinsic camera parameters of a hand-held video camera by tracking markers and natural features. Then some experimental results of 3-D reconstruction from real outdoor image sequences are demonstrated and evaluated to show the feasibility of the proposed method. Chapter 3 describes a method for dense depth estimation and integration of these dense data in a voxel space. The

---

\* Doctor's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT0161018, March 24, 2003.

reconstructed 3-D models of real outdoor scenes are also demonstrated. Finally, Chapter 4 summarizes the present study.

**Keywords:**

dynamic image analysis, camera parameter estimation, multi-baseline stereo, 3-D reconstruction

# 目次

1. 序論	1
1.1 現実環境の三次元モデル化に関する従来研究	1
1.1.1 能動的な三次元復元手法	2
1.1.2 受動的な三次元復元手法	6
1.2 本研究の位置づけと方針	12
1.3 本論文の構成	13
2. マーカと自然特徴点の追跡によるカメラパラメータの推定	14
2.1 序言	14
2.2 カメラパラメータ推定の概要	14
2.2.1 カメラパラメータの定義	15
2.2.2 カメラパラメータの推定手順	18
2.3 基準マーカの追跡	20
2.4 自然特徴点の追跡	20
2.4.1 Harris のインタレストオペレータを用いた自然特徴点候補 位置の検出	21
2.4.2 ロバスト推定を用いた自然特徴点の追跡	22
2.5 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	26
2.5.1 カメラの外部パラメータの推定	26
2.5.2 自然特徴点の三次元位置の推定	29
2.5.3 特徴点の信頼度の算出	32
2.5.4 自然特徴点の追加と削除	33
2.6 動画全体での最適化	34
2.7 屋外環境におけるカメラパラメータの復元実験	35
2.7.1 建物を対象としたカメラパラメータの復元 (実験 1)	37
2.7.2 町並みを対象としたカメラパラメータの復元 (実験 2)	42
2.8 評価実験	47
2.8.1 フレーム毎のカメラパラメータの推定精度の検証	47

2.8.2	カメラパラメータの累積的な誤差の評価	53
2.9	結言	58
<b>3.</b>	<b>拡張マルチベースラインステレオ法を用いた奥行き推定に基づく三次元モデル化</b>	<b>59</b>
3.1	序言	59
3.2	マルチベースラインステレオ法の概要	59
3.3	マルチベースラインステレオ法の拡張	61
3.3.1	低信頼度領域の検出	62
3.3.2	オクルージョンとスケールを考慮したマッチング尺度の算出	62
3.3.3	多重スケール法の利用	63
3.4	複数の動画像系列を用いたマルチベースラインステレオ	65
3.4.1	動画像間での日照条件の変化による影響の削減	65
3.4.2	奥行き探索に用いる画像の自動選択	66
3.5	ボクセル空間におけるモデルの復元	68
3.6	実験と考察	70
3.6.1	建物を対象とした三次元モデル化 (実験 1)	71
3.6.2	町並みを対象とした三次元モデル化 (実験 2)	81
3.6.3	考察	91
3.7	結言	92
<b>4.</b>	<b>結論</b>	<b>93</b>
4.1	まとめと考察	93
4.2	今後の展望	94
	謝辞	98
	参考文献	99
	付録	112

A. Tsai の手法によるカメラの内部パラメータのキャリブレーション	112
A.1 ピンホールカメラモデル	112
A.2 投影座標からデジタル画像への変換	114
A.3 Tsai によるカメラキャリブレーション手法	115
B. GUI ツールによるボクセルモデルの修正	116

## 目 次

1.1	奥行き画像の復元手法	2
1.2	三次元モデルの部分形状の復元	3
2.1	トータルステーション	15
2.2	カメラ座標系と世界座標系の関係	17
2.3	提案手法の処理の流れ	19
2.4	追跡の容易な自然特徴点の例	22
2.5	入力画像と Harris オペレータによる特徴量の出力結果	23
2.6	再投影の誤差	25
2.7	自然特徴点の三次元位置の初期値の算出	30
2.8	入力画像 (建物)	38
2.9	特徴点の追跡結果 (建物)	39
2.10	Match Move(建物)	40
2.11	カメラパラメータの推定結果 (建物)	41
2.12	入力画像 (町並み)	43
2.13	特徴点の追跡結果 (町並み)	44
2.14	Match Move(町並み)	45
2.15	カメラパラメータの推定結果 (町並み)	46
2.16	計算機シミュレーションによるフレーム毎のカメラ位置の推定誤差 (平面的なマーカ配置)	49
2.17	計算機シミュレーションによるフレーム毎のカメラ光軸方向の推定誤差 (平面的なマーカ配置)	50
2.18	計算機シミュレーションによるフレーム毎のカメラ位置の推定誤差 (空間的なマーカ配置)	51
2.19	計算機シミュレーションによるフレーム毎のカメラ光軸方向の推定誤差 (空間的なマーカ配置)	52
2.20	カメラ位置の推定誤差 (建物)	56
2.21	カメラ光軸の推定誤差 (建物)	56
2.22	カメラ位置の推定誤差 (町並み)	57

2.23	カメラ光軸の推定誤差 (町並み)	57
3.1	画素 $(x, y)$ の三次元位置と各画像上への投影直線	60
3.2	画像ピラミッドの例	64
3.3	奥行き探索に用いる画像の選択	67
3.4	画素 $(x, y)$ の奥行き推定値 $z$ によるボクセル空間への投票	69
3.5	入力画像 (建物)	74
3.6	復元されたカメラパラメータと自然特徴点の三次元位置 (建物)	75
3.7	SSD 算出のウィンドウサイズによる奥行き推定結果の比較に用いた入力フレームの画像 (建物)	76
3.8	SSD 算出のウィンドウサイズによる奥行き画像の比較 (建物)	76
3.9	奥行き画像の推定結果 (建物)	77
3.10	復元された三次元モデル (建物)	78
3.11	入力画像 (町並み)	83
3.12	復元されたカメラパラメータと自然特徴点の三次元位置 (町並み)	84
3.13	色の補正結果 (町並み)	85
3.14	SSD 算出のウィンドウサイズによる奥行き推定結果の比較に用いた入力フレームの画像 (町並み)	86
3.15	SSD 算出のウィンドウサイズによる奥行き画像の比較 (町並み)	86
3.16	奥行き画像の推定結果 (町並み)	87
3.17	復元された三次元モデル (町並み)	88
4.1	高解像度の全方位画像を取得可能な全方位センサ	96
4.2	光レーザに基づくレンジファインダによる計測データの可視化例	97
A.1	ピンホールカメラモデル	113
B.1	ボクセル修正ツールの構成	117
B.2	消しゴムツールの使用例	118
B.3	ツールによる修正後のモデル	119

## 表 目 次

2.1	カメラの内部・外部パラメータ一覧 . . . . .	16
2.2	実験に使用したカメラの内部パラメータ . . . . .	35
2.3	実験に使用したトータルステーションの仕様 (ノンプリズムモード)	35
2.4	実験に使用した自然特徴点追跡の閾値 . . . . .	36
2.5	累積的な誤差を含まないカメラパラメータの推定に用いた基準マ カの条件 . . . . .	53
2.6	累積的な誤差を含まないカメラパラメータのフレーム毎の推定誤 差の平均値 . . . . .	54
3.1	奥行き画像の復元とボクセル空間への統合における実験条件 . . . .	70



# 1. 序論

屋外環境の三次元モデルは、物体認識、景観シミュレーション、ナビゲーション、複合現実感など、多くの分野での利用が可能である。しかし、このような分野で用いられる三次元形状は、一般的にモデリングソフト [New02, Dis02] などを用いて人手によって作成されるために、多大なコストが必要とされ問題となる。一方、計算機やカーナビゲーションシステムの普及、バーチャルリアリティ技術の発展により、屋外環境の三次元モデルの需要は高まる一方である。このような背景から、屋外環境の三次元モデルを自動で獲得する研究は、現在強く開発が望まれている分野の一つである。

これに対し、古くはステレオ法に代表されるように、現実環境を観測位置に対する奥行き値の集合 (以下、奥行き画像) として表現し、復元する手法が数多く研究されてきた [YSK99, 井口 90]。奥行き画像を用いて完全な三次元モデルを得るためには、複数の観測位置から復元対象の部分的な形状 (以下、部分形状) を復元し統合を行う必要があるが、近年、これらの統合手法に関しても多数の研究が行われている [MSY96, BF97, Neu97, Pul99, ZS00, 増田 02, 西野 02]。しかし、従来提案されている手法では、それぞれ、復元範囲・精度やデータの取得にかかるコストなどの問題を残しており、屋外環境のような複雑で広域な環境の三次元モデルの自動作成は実現されていない。

以下では、まず、現実環境の三次元モデル化に関する従来手法を概観し、屋外環境の復元に対する応用の可能性について述べる。次に、従来手法の問題点、本研究の位置づけと研究方針を明確にする。

## 1.1 現実環境の三次元モデル化に関する従来研究

本節では、現実環境の三次元モデル化に関する従来手法を分類し詳述する。現実環境に対し、自動で三次元形状の復元を行う手法は、図 1.1 に示すように、撮影対象に積極的に電磁波を照射す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。

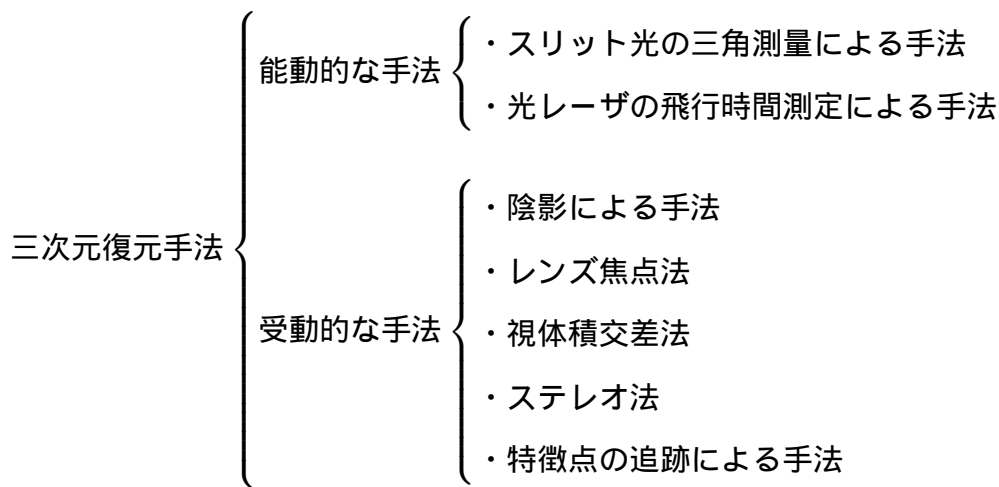


図 1.1 奥行き画像の復元手法

一般に、これらの手法を用いた三次元モデルの復元手法では、一度に三次元復元可能な範囲が限られているため、図 1.2 に示すように、現実環境に対して、部分的な三次元形状 (部分形状) しか得られない。このため、完全な三次元形状を得るためには部分形状の統合が必要となる。

以下では、現実環境の三次元モデル化に関する従来研究を部分形状の計測手法で分類し、それぞれの手法の特徴、屋外環境への応用の可能性および、部分形状の統合手法などについて述べる。

### 1.1.1 能動的な三次元復元手法

能動的な三次元復元手法は、レーザや投影光のような何らかのエネルギーを測定対象に照射し、三次元計測を行う手法である。能動的な三次元復元手法として、プロジェクタやレーザ光によって投影されるスリット光を近接する位置に設置した校正済みのカメラで測定し、三角測量の原理で奥行きを計測するスリット光の三角測量による手法と、計測対象に光パルスを照射し、反射して測定されるまでの時間 (time of flight) を計測する光レーザの飛行時間測定による手法がある。能動的な手法は、一般的に受動的な三次元復元手法に比べて測定の信頼度が高いとい

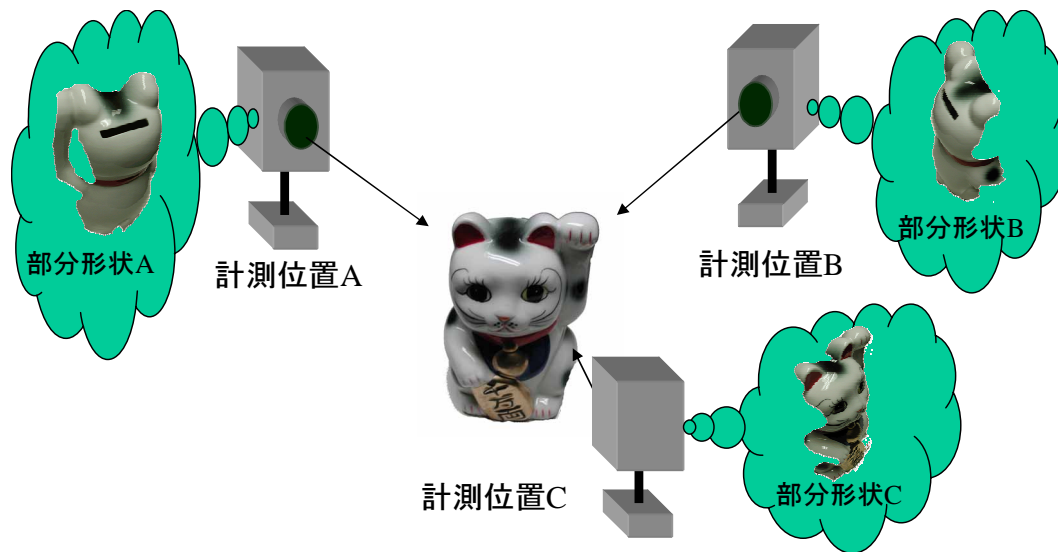


図 1.2 三次元モデルの部分形状の復元

う特長があるが、レーザや投影光を高精度に制御し測定する必要があるため機材が複雑となる。

#### スリット光の三角測量による手法

スリット光の三角測量に基づく手法として最も基本的なものは、スリット光を対象に照射し、近接して設置したカメラでスリット光の位置を検出することで三角測量の原理によって対象の部分形状を奥行き画像として復元する手法である [井口 90]。これは、後述する二眼ステレオ法の一方向のカメラをスリット光照射装置に置き換え、エピポラ線上の対応点探索を簡易化したものと考えられることができるため、アクティブステレオ法と呼ばれる。

この手法では、二眼ステレオにおける対応点のあいまいさが無いため、復元形状が安定するという特長がある。そのため、レーザによるスリット光を投影し、三角測量の原理を用いて計測を行うレンジファインダが、比較的小さな物体の計測機材として販売されている [MIN02, Cyb93]。

スリット光の照射を行う手法では、スリット光の照射位置を変化させながら画像を入力することが必要となるため、一枚の奥行き画像を得るためには多くの時間がかかるという問題がある。これに対して、奥行き画像の計測

時間を短縮するために同時に複数のスリット光を照射する手法や、二次元的な複数のパターンを変化させながら投影する手法が多数研究されている [佐藤 85, 盧 99, 塚本 00]. これらの手法は、基本的に数種類の二次元的なパターンを対象に投影することで計測が完了するため、計測時間を短縮することが可能である。ただし、投影パターンの周期性や物体面上の反射係数の違いによる観測パターンの色の変化によって、エピポーラ線上の探索に誤対応が発生する可能性が増すという問題があり、安定な復元を行うためには対象に制約が加わる。

スリット光の三角測量による手法によって復元された奥行き画像に対し、完全なモデルを得るためには、先に述べた部分形状の統合が必要となる。これに対し、レンジファインダをロボットアームなどに固定して移動させ、レンジファインダの位置・姿勢を制御する手法がある。回転式のレンジファインダ [Cyb93] では、観測台の上に物体を固定し、レンジファインダを物体の周囲で等速で回転させることで、物体の全周囲での奥行き画像とテクスチャ画像を計測する。この手法では、各画素を円筒面として展開すれば物体の側面形状の三次元モデルをテクスチャ付きで容易に作成可能であるが、物体上部、下部の形状が得られないという問題がある。しかし、小型の物体を計測するという目的においては実用的であるため、実際にこの装置を用いて作成された三次元モデルを用いて、物体の反射特性の推定に関する研究 [町田 01] や、顔の表情解析の研究 [辰野 96] などが行われている。

一方、部分形状を直接モデルの統合に用いる手法が多数提案されている [MSY96, BF97, Neu97, Pul99, 増田 02]. これらは、三次元空間中で部分形状同士のマッチングを行うことで形状の復元を行うために、部分形状の重複領域を多く持つことが重要となる。復元対象を小物体とした場合には、物体を回転させることで多数の部分形状を取得できるため、これらの手法を用いることで容易に精度の高い三次元モデルを得ることができる。

しかし、屋外環境の三次元モデル化という観点から見た場合、日照条件によってスリット光の観測が困難となる問題や、スリット光投影装置とカメラの間の距離を十分に確保できない問題があるため、精度の良い復元は困難である。

## 光レーザの飛行時間測定による手法

光レーザの飛行時間測定に基づくレンジファインダは、計測対象に光レーザパルス照射し、反射して測定されるまでの時間 (time of flight) を計測することでレンジデータを取得する。このような光レーザを二次元走査することで、レンジデータを奥行き画像として取得することが可能である。一般に、レーザの反射強度が一定以上あれば、高精度なレンジデータの獲得が可能であるという特長があり、屋外環境のような広域環境に対しても奥行き画像の計測を行うことが可能である。

このようなレンジファインダの開発に対しては、古くから研究がおこなわれてきた [NBD77, LJ77]。この方式ではテクスチャ情報が得られないという問題や、レーザ光の直視による失明の危険性などの問題があったが、近年直視しても問題のないアイセーフレーザが開発され、また、センサに 1 pixel CCD センサを内蔵しレンジデータと同期したカラー情報を獲得することで、各画素に対応したテクスチャを取得可能となっている。

また、非常に高価ではあるが、実際に面測定型 [Cyr00]、全周囲測定型 [Rie02] など複数のタイプの製品が販売されている。このタイプのレンジファインダは、屋内・屋外を問わず、レンジファインダから 100m 程度の範囲内であれば、安定で高精度なレンジデータを得ることができるという特長を持つ。一方、センサが点単位での二次元走査を行うために奥行き画像の取得に時間がかかり、移動を伴った計測は困難である。また、現在においてもレーザの反射率が低い濃色の物体の計測は困難である。

光レーザを用いたレンジファインダにおいても復元されるモデルは部分形状であるために、完全なモデルを作成するためには複数地点での計測データの統合が必要となる。これに対して、西野らは面測定型のレンジファインダを用いて文化財や遺跡の部分形状を多数の観測点から測定し、三次元空間内で点群データのマッチング処理を行うことで、三次元モデル化を行う手法を提案している [西野 02]。この手法では、復元対象に対して全周囲からの奥行き画像を得ることで、誤差の累積を防いでいる。また Zhao らは、屋外環境の三次元モデル化を目的として、全周囲測定型のレンジファインダを用いた復

元を行っている [ZS00]. この手法では, 奥行き画像上で二次元的な対応点探索を行い, 地表に対して鉛直上方向の回転と移動パラメータを求めている.

このような屋外環境を対象とした奥行き画像や三次元点群データのマッチングでは, 部分形状間で多くの共通領域が必要となるため, 多数の奥行き画像を獲得することが必要となる. 特に, 都市環境のように隠蔽が非常に多い環境においては, オクルージョンによるモデルの欠損がおこるために, 見た目に違和感の少ないモデルを作成するためには, 数メートル間隔での計測が必要となり, データの収集に手間がかかる. また, 機材が複雑なために非常に高価であるという問題や反射率の低い濃色の物体が復元されない問題がある. しかし, 今後, それらが解決されれば, 現実環境の三次元モデル化の有力な手法であるといえる.

### 1.1.2 受動的な三次元復元手法

受動的な三次元復元手法は, 一枚または複数枚の画像を入力として画像処理のみで三次元形状の復元を行う手法である. 陰影による手法, レンズ焦点法, 視体積交差法, ステレオ法, 特徴点の追跡による手法, などがある. 一般的に, 能動的な手法に対して, 機材の構成が簡易で安価となる点で有利であるが, 画像上で複数の候補からの推定処理を行うため, 精度の高い三次元モデルの作成には工夫が必要である.

#### 陰影による手法

陰影による手法は, 撮影対象の陰影画像と反射特性, 光源情報を用いて三次元復元を行う手法である. 光源情報と物体表面での反射率分布を既知として一枚の陰影画像から復元を行う手法 [Hor75] は, 画像濃度の勾配に基づいて物体表面の勾配を推定する. この手法では, 任意の点の奥行きを与えることで, 画像上の全ての点での奥行き画像を得ることができる. しかし, 事前に計測しておくべき情報が多く, また画像上のノイズに弱いという問題がある.

これらを拡張した手法として, 照度差ステレオ (photometric stereo) 法と呼

ばれる複数の陰影画像から復元を行う手法 [Woo80, Ike81] がある。照度差ステレオ法では、複数の光源位置に照明を設置し、カメラを固定してそれぞれの光源位置に対応した画像を取得して用いる。特に三点以上の光源位置を設定すれば、物体表面の勾配と反射率を同時に推定可能となるという特徴がある [Hor86]。また、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。

近年これらの手法の拡張として、陰影情報に加えて物体の輪郭情報を用いる手法 [戸所 98] や、また、人肌の三次元復元に関する研究などが行われている [MSO96]。しかし、陰影による手法では、暗室のように光源状況を任意に設定可能な環境でのみ利用が可能であり、また、カメラと物体の位置関係を固定することが前提であるため、部分形状の統合による三次元モデル化に関してはあまり議論がなされていない。また、実際に屋外環境の三次元モデル化の手法として用いることは難しいといえる。

#### 視体積交差法

視体積交差法は、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し、輪郭情報を空間に投影して得られる錐体 (視体積) の内側に対象が含まれるという仮定を用いて、物体の形状を取得する。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 [Zhe94]。また、任意のカメラ位置での撮影を可能とするために、ジャイロによって計測される姿勢情報を手がかりにモデルの統合を行う手法 [岡谷 00]、やマーカを用いてカメラ位置を推定し対象を復元する手法 [NW97] が提案されている。

これらの手法では、視体積の交差する領域を用いて復元を行うために、物体の凹み部分などの輪郭として現れない形状を復元することはできない。また、一般的に背景に暗幕などを用いることで輪郭の抽出を行うために、屋外環境のように輪郭情報を明確に得ることのできない環境に手法を適用することは難しい。

## ステレオ法

ステレオ法は、位置の異なる複数のカメラによって対象を撮影し、三角測量の原理を用いて三次元復元を行う手法である。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、画像上で各画素の対応を決定することによって、三角測量の原理により奥行き画像を推定する。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 [OK85] や、多重スケールを用いた手法 [横矢 93] 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 [MK85, AL87]、対応点のあいまいさを減少させる試みが成されている。さらに、複数台のカメラの投影中心を同一の平面上に配置し、特定のカメラの画像上で、各画素に対して奥行きを算出する、マルチベースラインステレオ法 [OK93] へと発展している。これらは、多数の画像を取得することにより、奥行き探索のあいまいさを減少させようというものである。

その他、現在においても三眼のカメラを搭載しリアルタイムで奥行き画像を推定するシステム [Poi01] や、二枚の航空写真から地図上の建造物の高さを推定し地図情報の自動更新を行う手法 [渡辺 02] など、ステレオ法とその応用分野に関する研究は盛んであり、ステレオ法は受動的な三次元復元手法として最も良く用いられる手法であるといえる。

一方、ステレオ法では、エピポーラ線の決定と正確な奥行き情報の算出のために、カメラ位置関係の算出が重要となり、多数のカメラキャリブレーション手法が研究されている。なかでも、代表的な手法として、三次元位置関係が既知のマーカを用いる手法がある [Tsa86, 出口 93]。これらの手法を用い、全てのカメラにマーカが写るような配置を行えば、カメラ間の位置関係が容易に算出でき、これらを用いて部分形状の統合が可能となる。実際に、Zhouらは数十台のカメラを用いて全方位からの奥行き画像計測を行うドームを構成し、マーカを用いて各カメラのキャリブレーションを行うことで、物体



をボクセル形状として復元している [ZG98].

屋外環境の復元を対象とした場合には、復元対象までの距離が屋内環境に比べて長くなるため、復元精度の低下を防ぐためには、カメラ間の距離 (ベースライン) を長くする必要がある。このような環境においては、複数のカメラに写るマーカを配置することが困難であるため、マーカを用いないカメラキャリブレーション手法を検討する必要がある。これに対して、GPS とジャイロを用いてカメラ位置・姿勢を推定し、ステレオ法を適用する手法 [横光 99, 松川 00] が提案されているが、GPS、ジャイロともに位置・姿勢の測定精度が十分ではなく、測定の誤差がそのまま復元されるモデルに影響するという問題がある。

### レンズ焦点法

レンズ焦点法 (Shape from Focus 法) 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 [Nay92, NS92, NN94]. この手法は、計測環境に対する事前知識や制約条件が必要ないという特長をもつ。しかし、画像のボケが発生する箇所を特定することが必要であるため、なめらかな形状のようなエッジを含まない領域を復元することは困難である。また、レンズの焦点を微小に変化させながら画像を計測する必要があるため、一枚の奥行き画像を得るために時間がかかるという問題がある。

同様の現象を用いた異なるアプローチとして、焦点距離の異なる二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 法と呼ばれる手法が研究されている [HM98, RC97]. Shape from Defocus 法では、二枚の画像のボケを解析することにより、各画素に対して焦点が合う奥行きを探索する。この手法を利用し、多焦点カメラを用いて一度に焦点の異なる複数の画像を取得することで、リアルタイムで奥行きを推定する手法に応用されている [NWN96].

Shape form Focus 法, Shape from Defocus 法は、ともに必要とされる事前情

報が少ないという利点があるが、被写界深度に幅があるため、屋外環境における高精度な三次元形状の獲得は困難である。また、画像上のノイズに弱いために、屋外環境において安定した推定を行うためには、多数の焦点画像を用いるなどの工夫が必要であるといえる。

#### 特徴点の追跡による手法

特徴点の追跡による手法は、複数の静止画像または動画像を入力とし、画像上の特徴点を複数の画像間に対応付ける（追跡する）ことで、特徴点の三次元位置とカメラの移動パラメータを復元する手法である。前述した他の手法とは異なり、任意の移動を伴った画像系列からの復元を行うところに特徴がある。なかでも、動画像を入力とする三次元復元の手法は、特徴点が画像上で連続的に移動することを利用して、特徴点の追跡を自動化できる。このため、動画像を入力とする三次元復元手法の研究が盛んに行われている。[TK92, SK94, BZM97, SY+99, 上原 00, PK+00, SM+02, 佐藤 02].

それらの代表的な手法に、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パラメータを同時に求める因子分解法 [TK92] がある。この手法は安定かつ高速に対象の復元を行うことができるため、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 [PK93, QK96, 出口 97, 植芝 97, 浮田 98, MK98]。しかし、カメラモデルを線形近似するためにカメラパラメータの復元精度に問題がある。さらに、動画像の全てのフレームに撮影されている特徴点を用いることを前提としているため、特徴点が全てのフレームに渡って連続して観測できない一般的な動画像からの三次元復元は困難である。

一方、射影的な復元を利用する手法が提案されている [BZM97, SY+99, RW00, PK+00]。これらは、各フレームにおいて、逐次的に射影的なカメラパラメータを復元し、射影座標系からユークリッド座標系への変換を行うことでカメラパラメータを復元する。しかし射影的な復元では、多くの自由度を残したまま逐次的な三次元復元を行うため、特徴点の誤対応による累積的な推定誤差が大きくなり、数十枚程度の画像からの復元を行うにとどまっている。

また一般に、動画像上の自然特徴点のみを用いる手法では、現実世界と復元されるデータとの位置関係およびスケールの情報が失われるため、カメラ間の相対位置関係を用いた部分形状の統合は困難である。加えて、上述した手法では部分形状が自然特徴点の間に面を構成する程度の簡易なモデルとして復元されており、これらを統合に用いることは難しい。

これらの問題を解決するため、動画像に加えてGPSやジャイロなどのセンサ類を利用する手法が提案されている。上原らは、動画像に加えてGPSと数値地図を用いることで町並みの三次元モデル化を試みている[上原00]。また、向井らはジャイロセンサを用いることで、カメラパラメータ推定の安定化を試みている[MO99]。しかし、GPSやジャイロセンサを直接位置データや姿勢データとして用いる場合には、一般的にGPSでは10m程度の位置推定誤差が、ジャイロセンサではジャイロのドリフトによる誤差の累積が避けられず、精度の良い復元は難しいといえる。

このように、特徴点の追跡による手法は、技術的な課題は残るが、復元対象と撮影位置・姿勢に対する制約が少なく、容易に多くの情報を獲得できるため、屋外環境において精度の良い三次元復元を行う手法として期待される手法の一つである。

## 1.2 本研究の位置づけと方針

前節で述べた各手法の特徴により、復元対象を限定せず自動で広域な環境の復元を行うためには、動画像を用いた特徴点の追跡による手法が適している。しかし、前述したように、動画像を用いた特徴点の追跡による手法を屋外環境で用いた場合、従来手法では以下の2つの問題がある。

### (1) カメラパラメータの推定に関する問題

特徴点の誤追跡によって、カメラパラメータの推定に誤差が累積し、復元精度が低下する。また、複数の動画像を用いて環境を撮影した場合には、動画像間のスケールや位置関係が不明となるため、部分形状の統合によって広域な環境を復元するというアプローチも困難である。

### (2) 三次元モデルの復元に関する問題

撮影対象に関して復元されるデータが自然特徴点の三次元位置のみであるため、復元されるモデルが特徴点の三次元位置の間に面を構成する程度の簡易なものとなり、細かな凹凸が復元されない。

本論文では、上記の問題を解決し、動画像から複雑な屋外環境の三次元モデル化を自動で行うことを目的とする。

本論文では、動画像を用いた屋外環境の三次元モデル化を行うにあたり、まず、自然特徴点に加えて、三次元位置が既知の特徴点(以下、基準マーカ)を画像上で追跡することで動画像のカメラの外部パラメータを精度良く推定する。次に、推定されたカメラパラメータを用い、マルチベースラインステレオ法によって各フレームでの奥行き画像を推定し、ボクセル空間への統合を行うことで三次元モデルの復元を行う。また、あらかじめ基準マーカの三次元座標系を統一しておくことで、複数の動画像を用いた場合においても、容易に部分形状の統合が可能となる。

### 1.3 本論文の構成

本論文の構成は以下のとおりである。2章では、特徴点として基準マーカと自然特徴点を用い、動画像撮影時のカメラパラメータを復元する手法について述べる。また、実際に屋外環境を撮影した動画像を入力として定量的な評価実験を行い、本手法の特徴と有用性について考察する。3章では、拡張マルチベースラインステレオ法による各フレームの奥行き情報の推定手法と、奥行き情報をボクセル空間で統合することで三次元モデルを復元する手法について述べる。実験では、カメラの移動パラメータが推定された動画像を用いて奥行き画像を推定し、三次元モデルの復元を行う。4章では、本論文のまとめと今後の課題を述べる。

## 2. マーカと自然特徴点の追跡によるカメラパラメータの推定

### 2.1 序言

現実環境を撮影した動画像からのカメラパラメータの推定は、三次元モデル復元、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。そのため動画像内の見かけの特徴点の動きから、特徴点の三次元位置とカメラパラメータを自動的に復元する手法 (shape from motion) に関する研究が盛んに行われてきた [TK92, PK93, SK94, BZM97, SY+99, PK+00, RW00, SK+01]。しかし、従来手法においては、手法上の制約からカメラパラメータの復元範囲が制限される問題や、自然特徴点の追跡誤差によってカメラパラメータの推定誤差が累積するという問題がある。

本章では、三次元位置が既知の特徴点 (以下、基準マーカ) と自然特徴点を画像上で自動追跡することで、上記の問題を解決する手法を提案する。本手法では、基準マーカの三次元位置を図 2.1 に示すようなトータルステーションと呼ばれる三次元測量機材を用いて計測し、初期フレームと最終フレームを含む少数の画像上で基準マーカを指定し、全体での最適化処理を行うことで、カメラパラメータの累積的な推定誤差を最小化する。また、追跡中の自然特徴点の三次元位置を逐次更新し、次のフレームでの特徴点の探索に利用することで、自然特徴点の安定した追跡を実現する。これにより、複雑で広域な環境を撮影した千枚程度から成る動画像のカメラパラメータを効率的かつ精度良く推定することが可能となる。

### 2.2 カメラパラメータ推定の概要

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、初期フレームおよび最終フレームを含む少数の画像上での基準マーカの画像座標を既知とし、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。



図 2.1 トータルステーション

以下では, 推定するカメラパラメータの定義と, カメラパラメータの推定手法の概要を述べる.

### 2.2.1 カメラパラメータの定義

本論文で用いるカメラの内部・外部パラメータの一覧を表 2.1 に示す. カメラの内部パラメータである, 焦点距離  $F$ , CCD のサイズ  $ccd_x, ccd_y$ , 画像サイズ  $scr_x, scr_y$ , 投影中心座標  $c_x, c_y$ , 画像スケール係数  $s_x$ , および, レンズ歪み率  $k_1, k_2$  は, 単一の動画像中では変化しないものとし, Tsai の手法 [Tsa86] などを用いて, あらかじめ計測しておくものとする (付録 A 参照). カメラの外部パラメータであるカメラの位置・姿勢は, 図 2.2 に示すように, 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  $(r_1, r_2, r_3)$  を, カメラの位置パラ

ルータとしてカメラ座標系におけるカメラの平行移動成分である  $t = (t_1, t_2, t_3)^T$  を用い、カメラの外部パラメータ  $M$  を以下のように定義する.

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

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

ただし,

$$\begin{aligned} s_1 &= \sin r_1, & s_2 &= \sin r_2, & s_3 &= \sin r_3 \\ c_1 &= \cos r_1, & c_2 &= \cos r_2, & c_3 &= \cos r_3 \end{aligned} \quad (3)$$

表 2.1 カメラの内部・外部パラメータ一覧

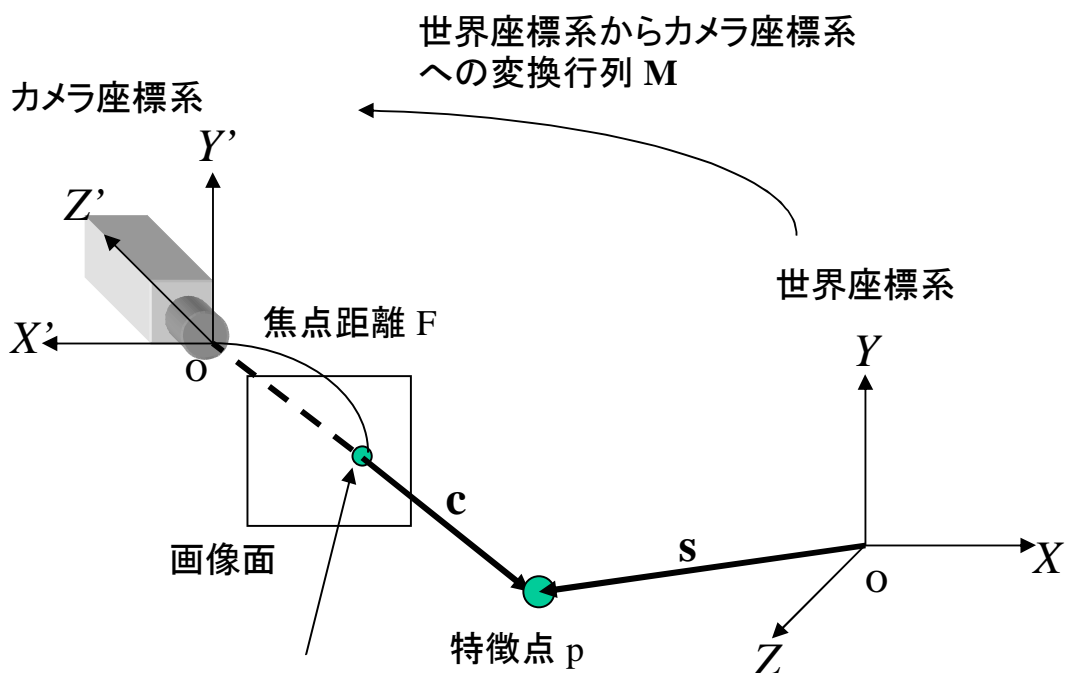
カメラの内部パラメータ

焦点距離 (mm)	F
CCD のサイズ (mm)	$ccd_x \times ccd_y$
画像サイズ (画素)	$scr_x \times scr_y$
投影中心座標 (画素)	$(c_x, c_y)$
スケール係数	$s_x$
レンズ歪み係数 ( $\text{mm}^{-2}, \text{mm}^{-4}$ )	$k_1, k_2$

カメラの外部パラメータ

姿勢パラメータ (radian)	$(r_1, r_2, r_3)$
平行移動成分 (mm)	$(t_1, t_2, t_3)$





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

図 2.2 カメラ座標系と世界座標系の関係

また、図 2.2 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  $p$  の三次元座標  $\mathbf{c}$  のデジタル画像上への投影座標  $\mathbf{x} = (x, y)$  は、以下のように定義される。

$$(x, y) = \left( -\frac{scr_x X'}{ccd_x Z'} F s_x + c_x, -\frac{scr_y Y'}{ccd_y Z'} F + c_y \right) \quad (5)$$

実際には、レンズ歪みの影響があるため、実際の入力画像上では、 $\mathbf{x} = (x, y)$  は  $\hat{\mathbf{x}} = (\hat{x}, \hat{y})$  に投影される。本手法では、以下の式を用いて、特徴点の入力画像上の座標  $(\hat{x}, \hat{y})$  をレンズ歪みを取り除いた座標  $\mathbf{x} = (x, y)$  に変換して用いる [Tsa86].

$$(x, y) = \left( (1 + k_1 r^2 + k_2 r^4)(\hat{x} - c_x) + c_x, (1 + k_1 r^2 + k_2 r^4)(\hat{y} - c_y) + c_y \right) \quad (6)$$

$$r = \sqrt{\left( \frac{ccd_x}{scr_x s_x} (\hat{x} - c_x) \right)^2 + \left( \frac{ccd_y}{scr_y} (\hat{y} - c_y) \right)^2} \quad (7)$$

### 2.2.2 カメラパラメータの推定手順

本節では、動画像中の特徴点を追跡し、カメラの外部パラメータを推定する手法について述べる。提案手法では図 2.3 に示すように、まず初期フレームにおいて画像上で三次元位置が既知の 6 個以上の基準マーカを画像上で指定する。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、2.5.1 項に述べる手法を用いて初期フレームのカメラの外部パラメータを推定する。次に、以下に示すフレーム毎の処理 (図中 A) を初期フレームから最終フレームまで繰り返すことにより、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逐次的に推定する。

- (a) マーカと自然特徴点の追跡: 基準マーカは色・形状の情報を用いて自動で追跡するか、あらかじめ手動により追跡する。自然特徴点は、Harris オペレータ [HS88] により追跡の容易な特徴点を検出して特徴点の候補位置とし、ロバスト推定 [栗田 00] によって誤追跡を検出しながら自動で追跡する。
- (b) カメラの外部パラメータの推定: 特徴点の画像上の座標と、その特徴点の三次元位置を画像上へ投影した座標との自乗距離を再投影の誤差として定義し、ステップ (a) で追跡された全て特徴点に関して信頼度による重みつきで再投影の誤差の和を最小化することでカメラパラメータを推定する。

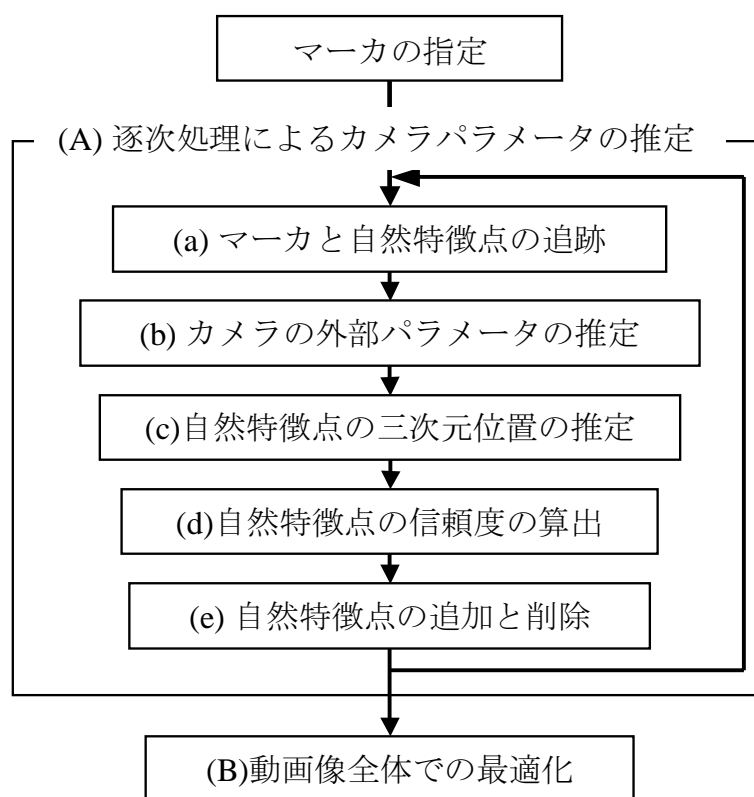


図 2.3 提案手法の処理の流れ

(c) 特徴点の三次元位置の推定: 自然特徴点が追跡された全てのフレームにおいて, 自然特徴点の再投影の誤差の和が最小となるような三次元位置を算出し, 自然特徴点の三次元位置を更新する.

(d) 自然特徴点の信頼度の算出: 自然特徴点の画像上の追跡誤差をガウス分布で近似することにより, 特徴点の信頼度を再投影の誤差の分散の逆数として定義し算出する.

(e) 自然特徴点の追加と削除: 特徴点の信頼度やテンプレート間の誤差などの複数の評価尺度を用いて, 自然特徴点の追加・削除を自動的に行う.

このような逐次処理においては, 自然特徴点の三次元位置を毎フレームで更新し, これらを用いて信頼度による重みつき評価関数を用いてカメラパラメータを

推定することで、基準マーカが画像上に存在しないフレームにおいても効率的かつ安定にカメラパラメータを復元することができる。しかし、この逐次処理によるカメラパラメータの推定には蓄積誤差が含まれるため、最後に、最終フレームを含む少数の画像上で基準マーカを指定し、推定されたカメラの外部パラメータを入力動画像全体で最適化する (図中 B)。

## 2.3 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・形状を用いた自動追跡または、手動による追跡を行う。基準マーカとして球状の単色マーカを用い、自動で追跡を行う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  $(f - 1)$  フレームのマーカ的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。手動での追跡を行う場合には、あらかじめ初期フレームを含む画像上で基準マーカの画像上の座標を指定しておく。

## 2.4 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 [TK91] や初期テンプレートを用いた自然特徴点の追跡手法 [藤井 00] では、以下の二つの問題によって特徴点追跡の誤差が累積する。

- (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ナなどの追跡の容易な点を Harris のインタレストオペレータ [HS88, SMB98] を用いて検出し、この出力が極大値となる点を自然特徴点の追跡位置の候補とする。また (b) に関

しては、既に推定されている自然特徴点の三次元位置とロバスト推定によって算出される暫定的なカメラパラメータを用いて、追跡すべき自然特徴点の探索範囲を限定することで解決する。

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。

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

本研究では、自然特徴点の追跡における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 [HS88] の出力が極大値を取る座標を自然特徴点の追跡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 2.4 の正方形窓で示すような、エッジのコナなどの、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 [SMB98] 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、同一の点に特徴点検出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  $\mathbf{x} = (x, y)$  の特徴量  $H(\mathbf{x})$  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  $\mathbf{W}$  において、画像上の輝度  $I$  の勾配  $I_x, I_y$  を用いて以下に示す行列  $\mathbf{A}$  を算出し、その行列の固有値  $\lambda_1, \lambda_2$  の極小値を求めることで得られる。

$$\mathbf{A} = \sum_{\mathbf{x} \in \mathbf{W}} \begin{pmatrix} I_x(\mathbf{x})^2 & I_x(\mathbf{x})I_y(\mathbf{x}) \\ I_x(\mathbf{x})I_y(\mathbf{x}) & I_y(\mathbf{x})^2 \end{pmatrix} \quad (8)$$

$$H(\mathbf{x}) = \min(\lambda_1, \lambda_2) \quad (9)$$

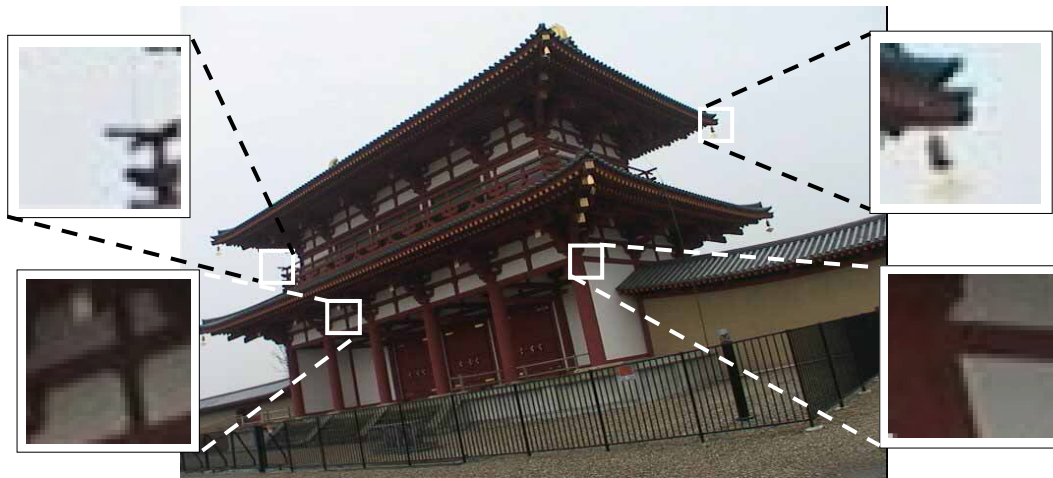


図 2.4 追跡の容易な自然特徴点の例

図 2.5(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  $H(x)$  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  $H(x)$  が極大値を取る位置を自然特徴点追跡の候補位置  $c = (c_0, c_1, \dots, c_n)$  とする.

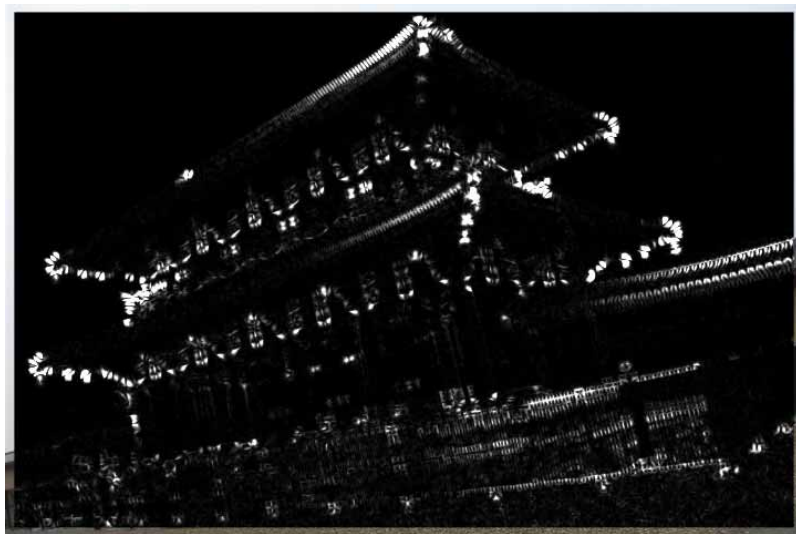
#### 2.4.2 ロバスト推定を用いた自然特徴点の追跡

前節の手法によって算出した自然特徴点の追跡の候補位置  $c = (c_0, c_1, \dots, c_n)$  に対し, テンプレートマッチングによって仮の特徴点追跡を行い, LMeds 基準を用いたロバスト推定 [栗田 00] 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. 次に, 算出される暫定的なカメラパラメータを用いて, 追跡すべき自然特徴点の探索範囲を限定することで, 特徴点追跡における誤追跡の問題を解決する. 以下に処理の流れを詳述する.

- (1) 第  $(f - 1)$  フレームで追跡されている各自然特徴点に対し, その特徴点の画像上の位置を中心とする探索窓内で SSD(Sum of Squared Differences) を評



(a) 入力画像



(b) 特徴量を輝度値に変換した画像

図 2.5 入力画像と Harris オペレータによる特徴量の出力結果

価尺度としたテンプレートマッチングを用いることで第  $f$  フレームにおける仮の追跡を行う。ただし、テンプレートマッチングは前述した Harris のインタレストオペレータ [HS88] の出力が極大値となる点に対してのみ行う。

- (2) 以下、ロバスト推定の繰り返し処理を開始する。  $i$  回目の繰り返し処理においては、まずステップ (1) で追跡された特徴点からランダムに  $n$  個  $P_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$  を選択し、次節に述べる手法によって仮のカメラパラメータ  $\hat{M}_i$  を推定する。
- (3) ステップ (1) で追跡された全ての特徴点に対して、図 2.6 に示すような、特徴点  $p$  が追跡された画像上の座標  $\mathbf{x}_{fp}$  と特徴点  $p$  の三次元座標  $S_p$  を  $\hat{M}_i$  によって画像上に投影した座標  $\hat{\mathbf{x}}_{fp}$  との距離の自乗誤差  $R_{ifp}$  (以下、再投影誤差) の中間値  $RM_i$  を算出する。ただし、ここでは  $S_p$  として第  $(f-1)$  フレームで推定されている三次元位置を用いる。なお、再投影誤差  $R_{ifp}$  とその中間値  $RM_i$  は以下の式によって表される。

$$R_{ifp} = |\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2 \quad (10)$$

$$RM_i = \text{med}(R_{if1}, R_{if2}, \dots, R_{ifm}) \quad (11)$$

ただし、 $m$  はステップ (1) で仮に追跡された特徴点の数である。

- (4) ステップ (2), (3) を  $g$  回繰り返して得られる複数の仮のカメラパラメータ  $(\hat{M}_1, \dots, \hat{M}_g)$  に対して、以下に示す LMedS 基準を満たすカメラパラメータを選択し、暫定カメラパラメータ  $\bar{M}_f$  とする。

$$LMedS = \min(RM_1, RM_2, \dots, RM_g) \quad (12)$$

- (5) 暫定カメラパラメータ  $\bar{M}_f$  を用いて第  $(f-1)$  フレームにおける全ての特徴点  $p$  の三次元座標  $S_p$  を画像上に投影し、その座標を中心とする探索窓内でステップ (1) と同様の対応づけを行い追跡位置を確定する。ただし、ここで用いる探索窓はステップ (1) で用いた探索窓より小さくする。



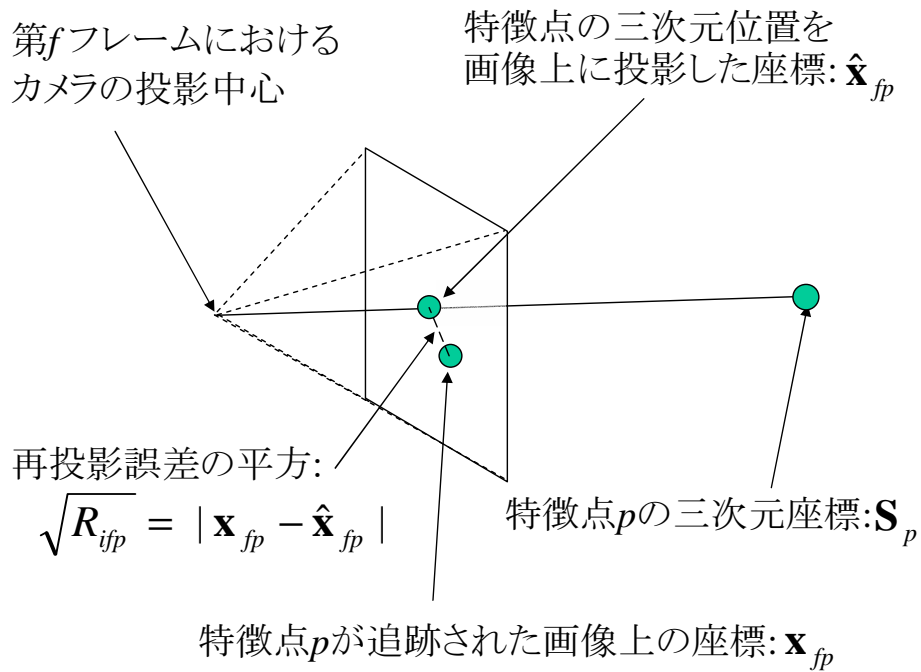


図 2.6 再投影の誤差

このような LMedS 基準を用いたロバスト推定において、誤追跡を含まない特徴点の集合のみで暫定カメラパラメータが推定される確率  $w$  は、以下の式により決定される。

$$w = 1 - (1 - (1 - \epsilon)^n)^g \quad (13)$$

ただし、 $\epsilon$  はステップ (1) で追跡される特徴点の中で、誤追跡が含まれる確率である。これにより、例えば、 $\epsilon = 0.1, n = 12$  とすれば、 $w = 0.9999$  となるためのサンプリング回数は、 $g = 28$  となる。

## 2.5 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

### 2.5.1 カメラの外部パラメータの推定

本項では、特徴点の画像上の二次元座標と、特徴点の世界座標における三次元座標を用いてカメラの外部パラメータを推定する。本手法においては、式(10)で示した再投影誤差  $R_{fp}$  を誤差尺度として利用する。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。

$$E_f = \sum_p W_p R_{fp} \quad (14)$$

ここで、 $W_p$  は特徴点  $p$  における重みであり、後述する特徴点  $p$  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ため、本手法では出口らの手法[出口 90]を利用し、線形最小自乗法によってカメラパラメータの初期値を算出し、補正を行った後に、非線形最適化を行う。以下にカメラの外部パラメータの推定手順を示す。

#### (1) 線形最小自乗法を用いたカメラの外部パラメータの初期値の算出

まず式(1)に示したカメラの外部パラメータの自由度を12として初期値を算出する。出口[出口 90]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  $p$  の単位焦点面での座標を  $(u_p, v_p)$ 、その世界座標系における同次座標を  $S_p = (X_p, Y_p, Z_p, 1)$  としたとき、以下の関係が成り立つ。

$$\begin{pmatrix} S_p & \mathbf{0} & -u_p S_p \\ \mathbf{0} & S_p & -v_p S_p \end{pmatrix} \mathbf{m} = \mathbf{0} \quad (15)$$

ただし、 $(u_p, v_p)$  は、入力画像面上の特徴点  $p$  のデジタル座標  $(x_p, y_p)$  に対して、カメラの内部パラメータを用いて以下の式で表される。

$$u_p = (x_p - c_x) \frac{ccd_x}{scr_x F s_x} \quad (16)$$

$$v_p = (y_p - c_y) \frac{ccd_y}{scr_y F} \quad (17)$$

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.

$$\begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} & -u_1\mathbf{S}_1 \\ \mathbf{S}_2 & \mathbf{0} & -u_2\mathbf{S}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{S}_n & \mathbf{0} & -u_n\mathbf{S}_n \\ \mathbf{0} & \mathbf{S}_1 & -v_1\mathbf{S}_1 \\ \mathbf{0} & \mathbf{S}_2 & -v_2\mathbf{S}_2 \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{S}_n & -v_n\mathbf{S}_n \end{pmatrix} \mathbf{m} = \mathbf{0} \quad (18)$$

$$\mathbf{W} = \begin{pmatrix} \sqrt{W_{f1}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \sqrt{W_{fn}} \end{pmatrix} \quad (19)$$

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 (18) を変形すると, 次式が得られる.

$$\mathbf{A}\mathbf{m}' = \mathbf{s} \quad (20)$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ \mathbf{S}_2 & \mathbf{0} & -u_2X_2 & -u_2Y_2 & -u_2Z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{S}_n & \mathbf{0} & -u_nX_n & -u_nY_n & -u_nZ_n \\ \mathbf{0} & \mathbf{S}_1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 \\ \mathbf{0} & \mathbf{S}_2 & -v_2X_2 & -v_2Y_2 & -v_2Z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{S}_n & -v_nX_n & -v_nY_n & -v_nZ_n \end{pmatrix} \quad (21)$$

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

これに対して線形最小自乗法を用いれば、擬似逆行列によって、 $\mathbf{m}' = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{s}$  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。

### (2) カメラパラメータの線形補正

ステップ (1) で初期値として推定したカメラパラメータに対して、自由度を 6 に補正する。ステップ (1) で用いた手法では、画像面に平行な軸上での誤差を最小化しているために、特徴点の画像面上での追跡誤差が、推定されるカメラの光軸方向に大きく影響する。このため、本手法では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。これによって、カメラパラメータを姿勢パラメータ  $(r_1, r_2, r_3)$  と平行移動成分である  $(t_1, t_2, t_3)$  の 6 パラメータで表現する。このとき、ステップ (1) で未定としたスケール係数も一意に決定される。

### (3) 非線形最適化による再投影誤差の最小化

ステップ (2) で求めたカメラパラメータを初期値とし、勾配法を用いて  $E_f$  を最小化することで各パラメータを最適化する。具体的には、以下の式を逐次的に

繰り返すことで各パラメータを更新し、式 (14) の  $E_f$  を最小化する。

$$r_1 \leftarrow r_1 - l_{r_1} \frac{\delta E_f}{\delta r_1}, \quad r_2 \leftarrow r_2 - l_{r_2} \frac{\delta E_f}{\delta r_2}, \quad r_3 \leftarrow r_3 - l_{r_3} \frac{\delta E_f}{\delta r_3} \quad (23)$$

$$t_1 \leftarrow t_1 - l_{t_1} \frac{\delta E_f}{\delta t_1}, \quad t_2 \leftarrow t_2 - l_{t_2} \frac{\delta E_f}{\delta t_2}, \quad t_3 \leftarrow t_3 - l_{t_3} \frac{\delta E_f}{\delta t_3} \quad (24)$$

ただし、矢印は更新を意味する。また、勾配法のスケール係数 ( $l_{r_1}, l_{r_2}, l_{r_3}, l_{t_1}, l_{t_2}, l_{t_3}$ ) は、各パラメータの更新時点で  $E_f$  が最小となるように直線探索を行い決定する。先に行った線形演算によって求めた初期値は真値に近い推定値であると考えられるため、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、再投影誤差の重み付きの和である  $E_f$  を最小化するカメラパラメータ  $M_f$  を求める。

### 2.5.2 自然特徴点の三次元位置の推定

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  $M_{(f-k)}, M_{(f-k+1)}, \dots, M_f$  から推定し、毎フレームで更新する。ここでは、カメラパラメータの推定と同様に、式 (10) で示した再投影誤差  $R_{fp}$  を誤差尺度として利用し、次の誤差関数  $E_p$  を勾配法によって最小化することで自然特徴点の三次元位置  $S_p$  を推定する。

$$E_p = \sum_f R_{fp} \quad (25)$$

$S_p$  の初期値は、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群との距離の自乗和が最小となる三次元位置を、線形最小自乗法によって算出する。以下に、式 (25) の最小化による自然特徴点の三次元位置の算出方法について詳述する。

#### (1) 線形最小自乗法による自然特徴点の三次元位置の初期値の算出

図 2.7 に示すように、第  $f$  フレームにおけるカメラの投影中心から、自然特徴点  $p$  の画像上の座標を空間に逆投影することで得られるベクトルの世界座標系における単位ベクトルを  $\mathbf{v}_{z_f} = (x_f, y_f, z_f)$  とする。まず  $\mathbf{v}_{z_f}$  に対して直交する単位ベ

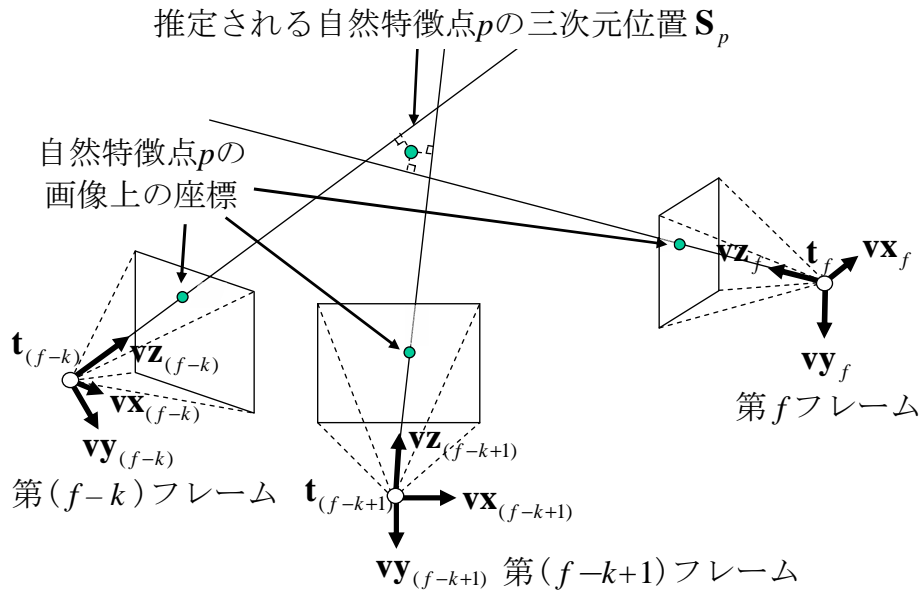


図 2.7 自然特徴点の三次元位置の初期値の算出

ベクトル  $vy_f$  を以下の式により決定する.

$$vy_f = \left( \frac{z_f}{\sqrt{x_f^2 + z_f^2}}, 0, -\frac{x_f}{\sqrt{x_f^2 + z_f^2}} \right) \quad (26)$$

また  $vy_f$  と  $vz_f$  のどちらにも直交する単位ベクトルを  $vx_f = vy_f \times vz_f$  とする. このとき, 第 $(f-k)$ フレームから第 $f$ フレームまでの各フレームの投影中心と自然特徴点  $p$  を結ぶ各直線と, 自然特徴点  $p$  の三次元位置  $S_p$  の距離の自乗和  $R$  は,  $vx_f, vy_f$  および, 世界座標におけるカメラ位置  $t_f = (tx_f, ty_f, tz_f)$  を用いて以下の式により算出できる.

$$R = \sum_{i=f-k}^f (rx_i^2 + ry_i^2) \quad (27)$$

$$\mathbf{r} = \begin{pmatrix} rx_{(f-k)} \\ rx_{(f-k+1)} \\ \vdots \\ rx_f \\ ry_{(f-k)} \\ ry_{(f-k+1)} \\ \vdots \\ ry_f \end{pmatrix} = \begin{pmatrix} \mathbf{vX}_{(f-k)} & tx_{(f-k)} \\ \mathbf{vX}_{(f-k+1)} & tx_{(f-k+1)} \\ \vdots & \vdots \\ \mathbf{vX}_f & tx_f \\ \mathbf{vY}_{(f-k)} & ty_{(f-k)} \\ \mathbf{vY}_{(f-k+1)} & ty_{(f-k+1)} \\ \vdots & \vdots \\ \mathbf{vY}_f & ty_f \end{pmatrix} \begin{pmatrix} X_p \\ Y_p \\ Z_p \\ 1 \end{pmatrix} \quad (28)$$

これを变形すると、次式が得られる。

$$\mathbf{r} + \mathbf{t} = \mathbf{C}\mathbf{S}_p \quad (29)$$

$$\mathbf{t} = \begin{pmatrix} -tx_{(f-k)} \\ -tx_{(f-k+1)} \\ \vdots \\ -tx_f \\ -ty_{(f-k)} \\ -ty_{(f-k+1)} \\ \vdots \\ -ty_f \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{vX}_{(f-k)} \\ \mathbf{vX}_{(f-k+1)} \\ \vdots \\ \mathbf{vX}_f \\ \mathbf{vY}_{(f-k)} \\ \mathbf{vY}_{(f-k+1)} \\ \vdots \\ \mathbf{vY}_f \end{pmatrix} \quad (30)$$

これに対して線形最小自乗法を用いれば、 $R$ を最小とする三次元座標  $\mathbf{S}_p = (X_p, Y_p, Z_p)$  は、擬似逆行列によって、 $\mathbf{S}_p = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{t}$  により得られる。

## (2) 非線形最適化による再投影誤差の最小化

カメラパラメータの算出に用いた手法と同様に、以下の式を逐次的に繰り返すことで再投影誤差の和  $E_p$  を最小化し、自然特徴点の三次元位置を更新する。

$$X_p \leftarrow X_p - l_{X_p} \frac{\delta E_p}{\delta X_p}, \quad Y_p \leftarrow Y_p - l_{Y_p} \frac{\delta E_p}{\delta Y_p}, \quad Z_p \leftarrow Z_p - l_{Z_p} \frac{\delta E_p}{\delta Z_p} \quad (31)$$

スケール係数  $(l_{X_p}, l_{Y_p}, l_{Z_p})$  は、各パラメータの更新時点で  $E_p$  が最小となるように直線探索を行い決定する。これにより、再投影誤差の和である  $E_p$  を最小化する自然特徴点の三次元位置  $\mathbf{S}_p = (X_p, Y_p, Z_p)$  を求める。

### 2.5.3 特徴点の信頼度の算出

第  $f$  フレームにおいて算出される自然特徴点の信頼度は，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，第  $f$  フレームにおける自然特徴点の追加・削除の指標として用いられる．本項では第  $f$  フレームにおける自然特徴点の信頼度の算出について述べる．

先に求めた第  $f$  フレームにおける特徴点  $p$  の画像上の座標  $\mathbf{x}_{fp}$  は，追跡誤差により一般的に，特徴点  $p$  の三次元位置  $S_p$  を画像上に投影した座標  $\hat{\mathbf{x}}_{fp}$  と一致しない．そこで，この追跡誤差が  $\hat{\mathbf{x}}_{fp}$  周辺に分散  $\sigma_p^2$  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， $\mathbf{x}_{fp}$  が真値である確率  $p(\mathbf{x}_{fp})$  は次式のようになる．

$$p(\mathbf{x}_{fp}) = \frac{1}{2\pi\sigma_p^2} \exp\left(-\frac{|\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2}{2\sigma_p^2}\right) \quad (32)$$

これを用い，第  $f$  フレームにおける全ての特徴点  $p$  の確率密度関数の積  $P_f$  を考える．

$$P_f = \prod_p p(\mathbf{x}_{fp}) \quad (33)$$

このような確率密度関数の積  $P_f$  を最大とするカメラパラメータ  $M_f$  は，

$$EM_f = \sum_p \log(2\pi\sigma_p^2) + \sum_p \frac{|\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2}{2\sigma_p^2} \quad (34)$$

を最小化することにより得られる．ただし，自然特徴点の追跡誤差の分散  $\sigma_p^2$  は，第  $f$  フレームのカメラパラメータ  $M_f$  に対して独立であるため，式 (34) の右辺の第一項は定数となり，実際には以下の関数を最小化すればよい．

$$EM_f = \sum_p \frac{|\mathbf{x}_{fp} - \hat{\mathbf{x}}_{fp}|^2}{2\sigma_p^2} \quad (35)$$

本手法では， $\sigma_p^2$  として再投影誤差の分散を用い，式 (14) と式 (35) 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  $p$  の信頼度  $W_p$  を次式のように定義し，毎フレームで更新する．

$$W_p = \frac{1}{2\sigma_p^2} = \frac{k+1}{2} \left\{ \sum_{i=f-k}^f |\mathbf{x}_{ip} - \hat{\mathbf{x}}_{ip}|^2 \right\}^{-1} \quad (36)$$



#### 2.5.4 自然特徴点の追加と削除

自然特徴点の追加は, 2.4.1 項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群の成す最大角が一定値以上

また, 以下のいずれかの条件を満たす自然特徴点を削除する.

- 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
- 再投影誤差が一定値以上

## 2.6 動画像全体での最適化

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、逐次的な復元処理で用いた基準マーカに加えて、最終フレームを含む少数のフレームの画像上で基準マーカを指定し、全体での最適化を行うことで、累積的な誤差を最小化する。ここでは、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、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。

$$E = \sum_f \left( A_f \sum_p W_p |x_{fp} - \hat{x}_{fp}|^2 \right) \quad (37)$$

ここで  $W_p$  は、逐次処理で得た自然特徴点  $p$  の信頼度である。 $A_f$  は、フレームの信頼度による重みであり、ここでは、 $A_f$  として基準マーカが存在しないフレームでは1を、基準マーカが存在するフレームでは1に対して非常に大きな値  $C$  を設定する。これによって、マーカが存在するフレームでの推定結果を真値とみなして、動画像全体の再投影誤差を最小化することができる。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収束しやすくなり、また高速に求めることができる。

## 2.7 屋外環境におけるカメラパラメータの復元実験

提案手法の有効性を確認するため、実環境における建物として奈良・朱雀門を、広域で複雑な環境として大阪・住道の町並みをビデオカメラで撮影し、カメラパラメータの復元実験と定量的な評価を行った。両実験ともに、ワイドレンズ (Sony VCL-HG0758) を取り付けたビデオカメラ (Sony DSR-DP-150) を用いて撮影を行い、得られた動画像を入力とした。ただし、カメラの内部パラメータは Tsai の手法 [Tsa86] によりあらかじめ算出した。また、基準マーカは三次元測量機器トータルステーション (Leica TCR1105XR) を用いてノンプリズムモード (プリズムなし) で計測した。

実験で使用したカメラの内部パラメータを表 2.2 に、トータルステーションの仕様を表 2.3 に、自然特徴点の追跡の閾値を表 2.4 に示す。各実験においては、撮影方法の違いによって、自然特徴点のフレームイン・フレームアウトの頻度および自然特徴点の画像上の動きに大きな違いが生じるため、自然特徴点の追跡の閾値を経験的に調整した。

表 2.2 実験に使用したカメラの内部パラメータ

焦点距離 (mm)	4.093326
CCD 画素サイズ (mm)	4.4 × 3.3
画像サイズ (画素)	720 × 480
投影中心座標 (画素)	( -0.8, 6.5 )
スケール係数	0.975541
レンズ歪み係数 $k_1$ (mm <sup>-2</sup> )	0.01476244
レンズ歪み係数 $k_2$ (mm <sup>-4</sup> )	-0.0002585938

表 2.3 実験に使用したトータルステーションの仕様 (ノンプリズムモード)

型式	Leica TCR1105XR
測角精度	5"
測定精度	3mm+2ppm
測定時間	3 秒程度
測定範囲	200m

表 2.4 実験に使用した自然特徴点追跡の閾値

自然特徴点の追跡のパラメータ

	建物を対象 とした実験	町並みを対象 とした実験
テンプレートのサイズ (画素)	13×13	9×9
仮追跡の探索範囲 (画素)	61×61	71×51
限定した探索範囲 (画素)	9×9	9×9

自然特徴点の追加の閾値

	建物を対象 とした実験	町並みを対象 とした実験
信頼度の最小値 (分散の逆数)	$\frac{1}{6}$	$\frac{1}{3}$
フレーム間のテンプレートの最低相関値	0.97	0.98
投影中心と点を結ぶ直線間の成す最小角 (度)	2.5	6.0
最低追跡フレーム数 (枚)	20	12

自然特徴点の削除の閾値

	建物を対象 とした実験	町並みを対象 とした実験
信頼度の最小値 (分散の逆数)	$\frac{1}{8}$	$\frac{1}{12}$
フレーム間のテンプレートの最低相関値	0.97	0.98
再投影誤差の平方の最大値 (画素)	3.0	3.0

### 2.7.1 建物を対象としたカメラパラメータの復元 (実験 1)

本実験においては、手持ちのビデオカメラを用いて朱雀門を中心として円弧を描くように撮影を行い、図 2.8 を含む画像 982 枚 (720×480 画素、プログレッシブ撮影) を得た。また、あらかじめ、図 2.9 の第 1 フレームにおいて 印で示す点を含む複数の自然特徴点を基準マーカとし、その三次元位置関係を、トータルステーション (Leica TCR1105XR) を用いて計測した。基準マーカの画像上の位置は初期フレームから第 100 フレーム、第 480 フレーム、最終フレームの合計 102 枚の画像上で手動で指定した。

図 2.9 に特徴点の追跡結果を示す。図中の 印は指定した基準マーカを、×印は追跡された自然特徴点を表している。同図に示すように、多数の自然特徴点が追加・削除を伴って安定に追跡された。本実験においては、各フレームで平均 190 点の自然特徴点がそれぞれ平均 100 フレーム間自動で追跡された。実験には、PC(Pentium4 2GHz, Memory 1GB) を用い、カメラパラメータの復元処理に要した時間は、自然特徴点の自動追跡を含む逐次的な復元処理に約 47 分、全体での最適化処理における 100 回の繰り返し計算に約 79 分であり、全体最適化後の自然特徴点の再投影誤差の平均値の平方は 0.95 画素であった。

図 2.10 は、推定されたカメラパラメータを用いて、入力画像上に CG で描かれる人物を重畳表示したものである。実験から、動画像中で CG による人物が一貫して同じ位置に表示されていることが確認できた。このため、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ことが分かる。

図 2.11 に復元された自然特徴点の三次元位置とカメラパラメータの推定結果を示す。曲線は推定されたカメラパスを、錘台は 50 フレーム毎のカメラの姿勢を表す。本実験において復元されたカメラパスの長さは約 76m であった。同図から分かるように、カメラパス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。



第1 フレーム



第560 フレーム



第140 フレーム



第700 フレーム



第280 フレーム



第840 フレーム



第420 フレーム



第982 フレーム

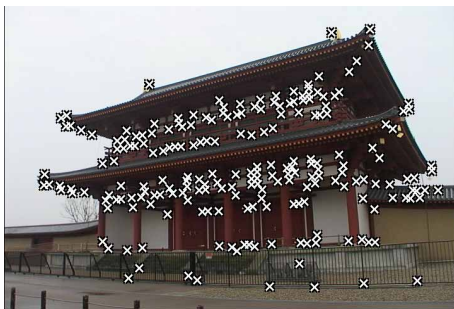
図 2.8 入力画像 (建物)



第 1 フレーム



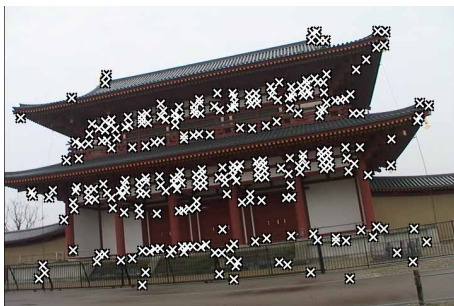
第 560 フレーム



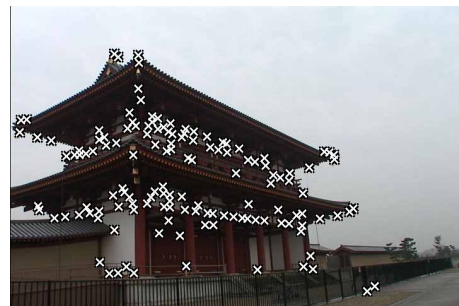
第 140 フレーム



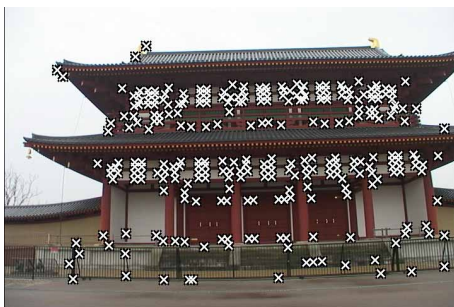
第 700 フレーム



第 280 フレーム



第 840 フレーム



第 420 フレーム



第 982 フレーム

図 2.9 特徴点の追跡結果 (建物)





第 1 フレーム



第 560 フレーム



第 140 フレーム



第 700 フレーム



第 280 フレーム



第 840 フレーム



第 420 フレーム



第 982 フレーム

図 2.10 Match Move(建物)



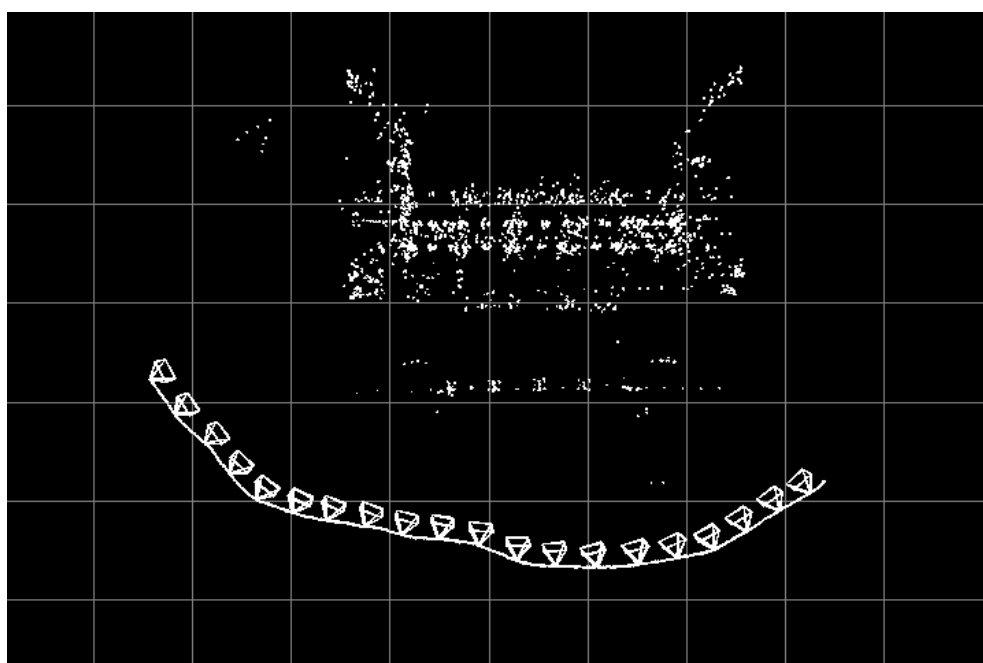
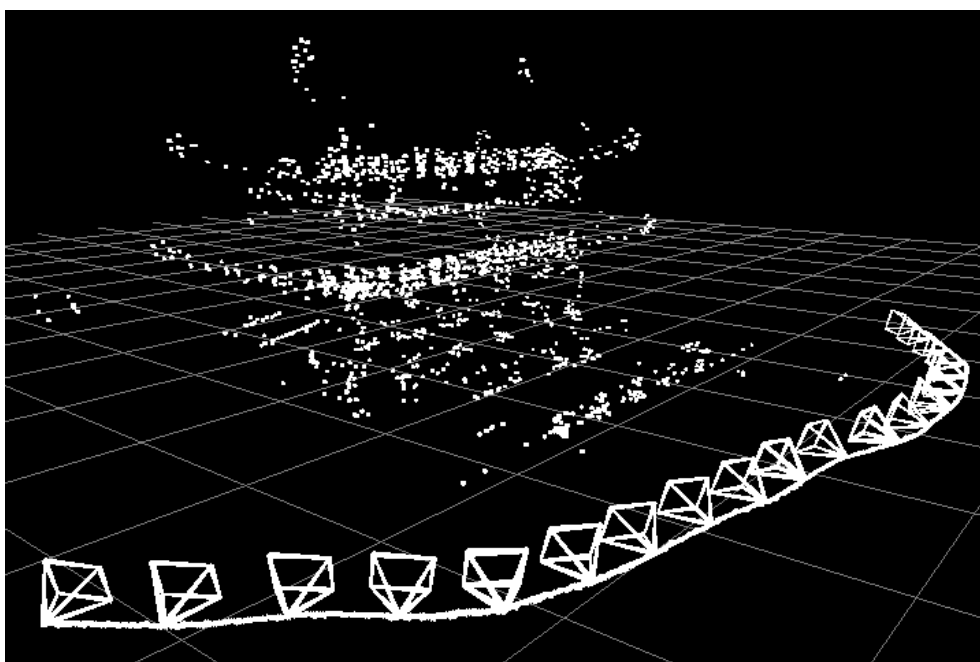


図 2.11 カメラパラメータの推定結果 (建物)

### 2.7.2 町並みを対象としたカメラパラメータの復元 (実験 2)

本実験では、ビデオカメラを車の上に三脚で固定し、住道の町並みを撮影しながら走行することで、図 2.12 を含む画像 500 枚 (720×480 画素、プログレッシブ撮影) を得た。ただし、特徴点の頻繁なフレームアウトが起こることを防ぐために、画像系列を逆再生して利用した。また、前節の建物の復元実験と同様に、図 2.13 の第 1 フレームに 印で示す自然特徴点を基準マーカとし、三次元位置関係をトータルステーションを用いて座標系を統一して計測した。基準マーカの画像上の位置は、初期フレームから第 50 フレーム、第 250 フレーム、最終フレームの合計 52 枚の画像上で指定した。

図 2.13 に特徴点の追跡結果を示す。図中の 印は指定した基準マーカを、×印は追跡された自然特徴点を表している。本実験において追跡された自然特徴点の数は各フレーム平均 110 点であり、それぞれ平均 20 フレーム間自動で追跡された。また、実験には前項と同じ計算機を用い、カメラパラメータの復元処理に要した時間は、自然特徴点の自動追跡を含む逐次的な復元処理に約 17 分、全体での最適化処理における 100 回の繰り返し計算に約 30 分であり、全体最適化後の自然特徴点の再投影誤差の平均値の平方は、0.82 画素であった。

図 2.14 は、前項の建物を対象とした実験と同様に、推定されたカメラパラメータを用いて、入力画像上に CG で描かれる鉄塔を重畳表示したものである。ただし、この実験では隠蔽関係を再現するために、色の閾値処理によって空を検出し、空の領域にのみ CG を描画している。動画像中で CG による鉄塔は一貫して同じ位置に表示されていることが確認できるため、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ことが分かる。

図 2.15 に復元された自然特徴点の三次元位置とカメラパラメータの推定結果を示す。前項と同様に、曲線は推定されたカメラパスを、錘台は 50 フレーム毎のカメラの姿勢を表す。本実験において復元されたカメラパスの長さは約 130m であった。同図から分かるように、オクルージョンの発生する複雑な町並みを撮影した動画像を入力として用いた場合にも、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。



第 1 フレーム



第 280 フレーム



第 70 フレーム



第 350 フレーム



第 140 フレーム



第 420 フレーム



第 210 フレーム

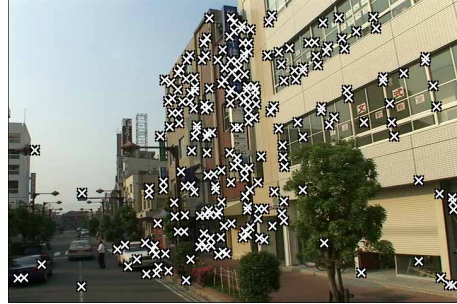


第 500 フレーム

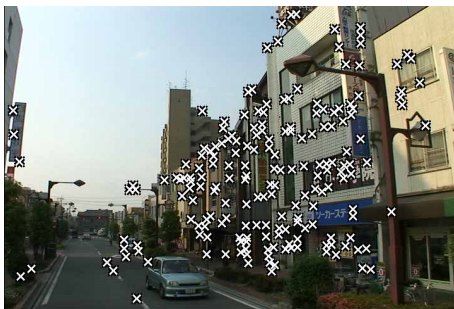
図 2.12 入力画像 (町並み)



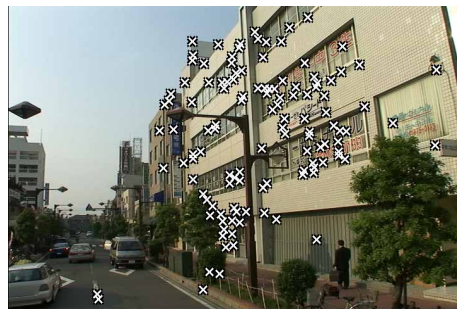
第 1 フレーム



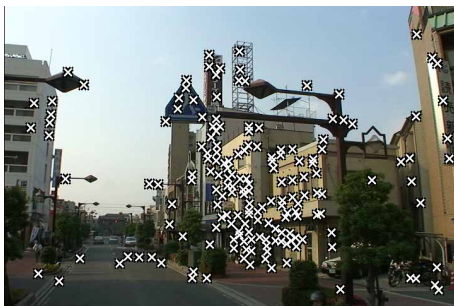
第 280 フレーム



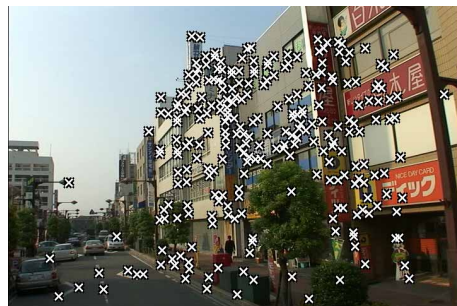
第 70 フレーム



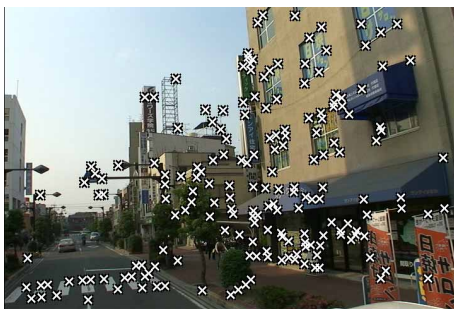
第 350 フレーム



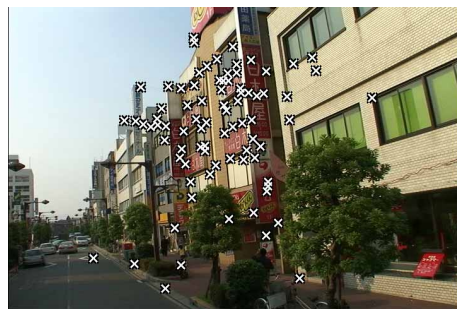
第 140 フレーム



第 420 フレーム



第 210 フレーム



第 500 フレーム

図 2.13 特徴点の追跡結果 (町並み)





第 1 フレーム



第 280 フレーム



第 70 フレーム



第 350 フレーム



第 140 フレーム



第 420 フレーム



第 210 フレーム



第 500 フレーム

図 2.14 Match Move(町並み)

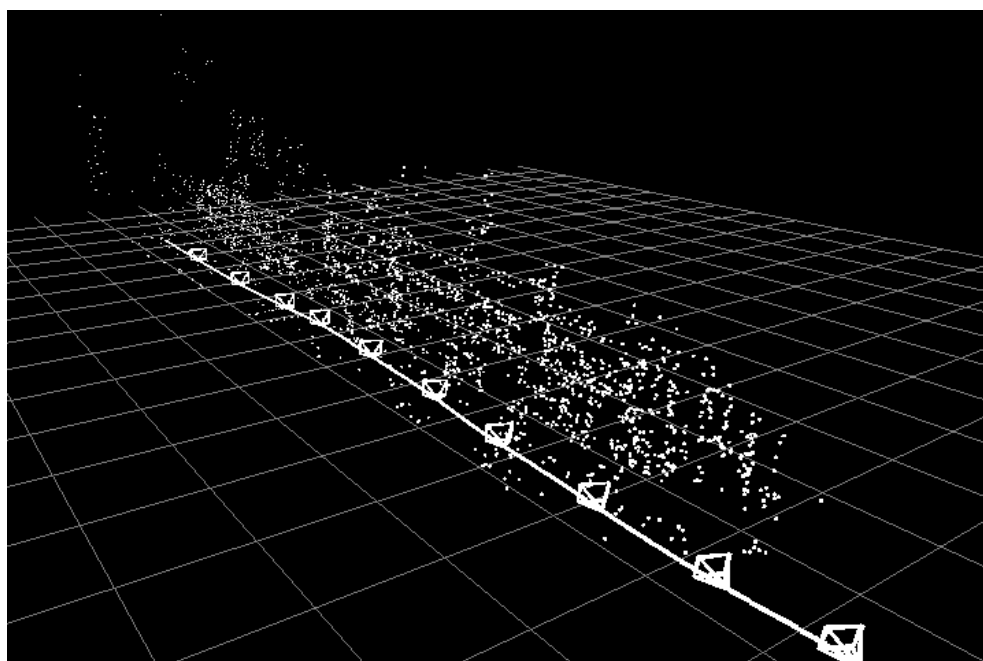
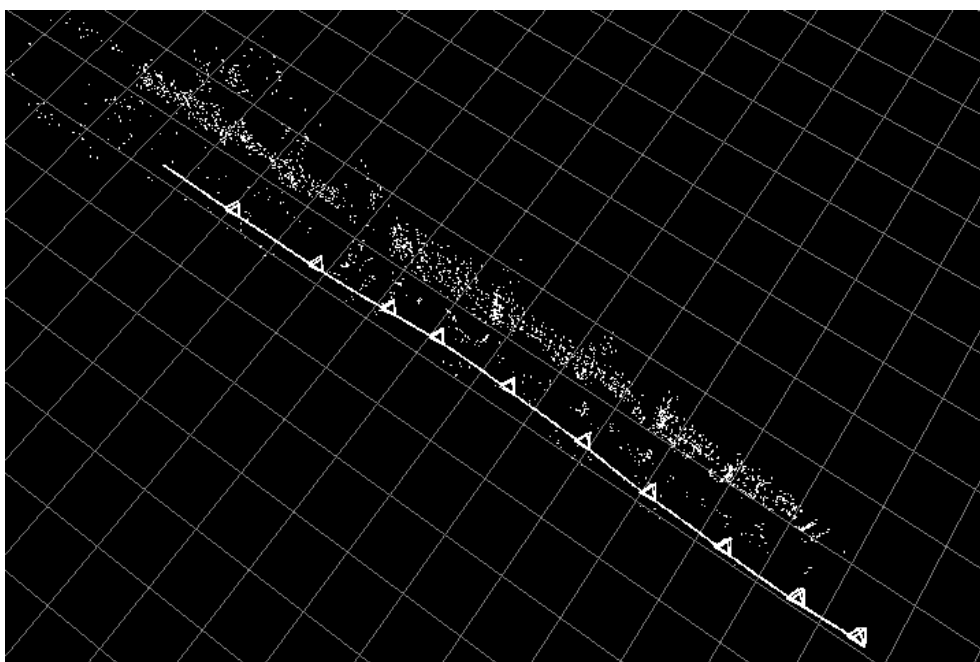


図 2.15 カメラパラメータの推定結果 (町並み)

## 2.8 評価実験

本節では、前述したカメラパラメータの推定手法に関する精度評価を行なう。以下では、まず各フレームにおいて発生する画像上の追跡誤差および量子化誤差に起因するカメラパラメータの推定誤差について、計算機シミュレーションによる検証を行い、続いてカメラパラメータの累積的な誤差について、前節で述べた実験 1, 2 の結果を評価する。

### 2.8.1 フレーム毎のカメラパラメータの推定精度の検証

ここでは、各フレームにおいて特徴点の追跡誤差がカメラパラメータの推定精度に与える影響を検証する。以下では、特徴点の画像への投影座標に対して、次に示す (1), (2) の誤差が画像面上の各軸方向に発生すると仮定する。

- (1) 追跡座標の量子化に起因する  $[-0.5, 0.5]$  の範囲を持つ一様分布の誤差
- (2) 画像上のノイズや誤追跡による標準偏差  $\sigma$  のガウス分布をもつ追跡誤差

本手法においては、透視投影の非線形のカメラモデルを用い再投影誤差の和を最小化することで、カメラパラメータの推定を行なっているため、(1), (2) の誤差とカメラパラメータの推定精度の関係について、確率論的な解析を行なうことは困難である。そこで、これらの誤差を計算機シミュレーションによって解析する。

計算機シミュレーションにおいては、仮想空間内においてカメラパラメータの真値と特徴点の三次元位置関係を既知とし、マーカをカメラから一定の距離の平面上に一様に分布させた。また、既知である特徴点の三次元位置を、実験 1, 2 で用いたカメラの内部パラメータを用いて画像面上に投影し、(1), (2) の誤差を擬似的に発生させることで、特徴点の画像上の座標とした。このような条件の下で、マーカとカメラ間の距離  $le$ 、画像に投影されるマーカの数  $ne$ 、(2) の誤差の標準偏差  $\sigma$  を変化させ、2.5.1 項に述べたカメラパラメータの推定手法を用いることで、真値との誤差を算出した。ただし、空間中の特徴点の配置を完全な平面とした場合には、2.5.1 項の手法における線形演算による初期値の算出が不可能となるため、特徴点の三次元位置の真値に対して、奥行き方向に  $le/1000$  以下のばらつきを与え

である。また、本実験においては  $(le, ne, \sigma)$  の各組に対して、100 回ずつカメラパラメータを推定し誤差の平均値の算出を行なっている。

図 2.16 にシミュレーションによって算出されたカメラ位置の誤差をグラフで示す。また、同図中に、標準偏差  $\sigma$  に対応する再投影誤差の平均値の平方  $R$  を画素単位で示す。同図から、標準偏差  $\sigma$  によらず、マーカの数  $ne$  が増加するに従ってカメラ位置の推定誤差が減少しており、カメラとマーカの距離  $le$  に比例して誤差が増加していることが確認できる。図 2.17 に算出されたカメラの光軸方向の誤差の平均値を示す。カメラ位置と同様に、マーカの数  $ne$  が増加するに従って推定誤差が減少しているが、光軸方向の誤差はカメラとマーカの距離  $le$  には依存していない。また、両図から、特徴点の追跡誤差の標準偏差  $\sigma$  の増加に従って、カメラ位置・光軸方向の推定誤差が共に増大していることが分かる。

続いて、空間的なマーカの配置についても検討する。ここでは、マーカがカメラの光軸方向に対して奥行き  $le=20000\text{mm}$  を中心とする一定の範囲内  $[-he \times le, he \times le]$  で一様に分布すると仮定し、 $(he, ne, \sigma)$  の各組に対して、前述した実験と同様の計算機シミュレーションを行なった。図 2.18 にシミュレーションによって算出されたカメラ位置の誤差の平均値を、図 2.19 に算出されたカメラの光軸方向の誤差の平均値を示す。これらのグラフから、マーカの分布幅  $(2 \times he \times le)$  を大きくとれば、 $(\sigma, ne)$  にかかわらず、カメラパラメータの推定精度が向上することが確認できる。特に、特徴点の追跡誤差が大きくマーカの数が少ない場合には、空間的なマーカの配置を行なうことで、カメラ位置および光軸方向の推定精度が大幅に向上することが確認できる。

これらのシミュレーション結果から、各フレームにおけるカメラパラメータの推定では、以下の点を考慮することが、カメラパラメータの推定精度の向上に有効であるといえる。

- (a) 特徴点を空間的に配置すること
- (b) 特徴点の追跡精度を向上させること
- (c) 推定に用いる特徴点の数を増やすこと
- (d) 撮影対象までの距離を短くすること



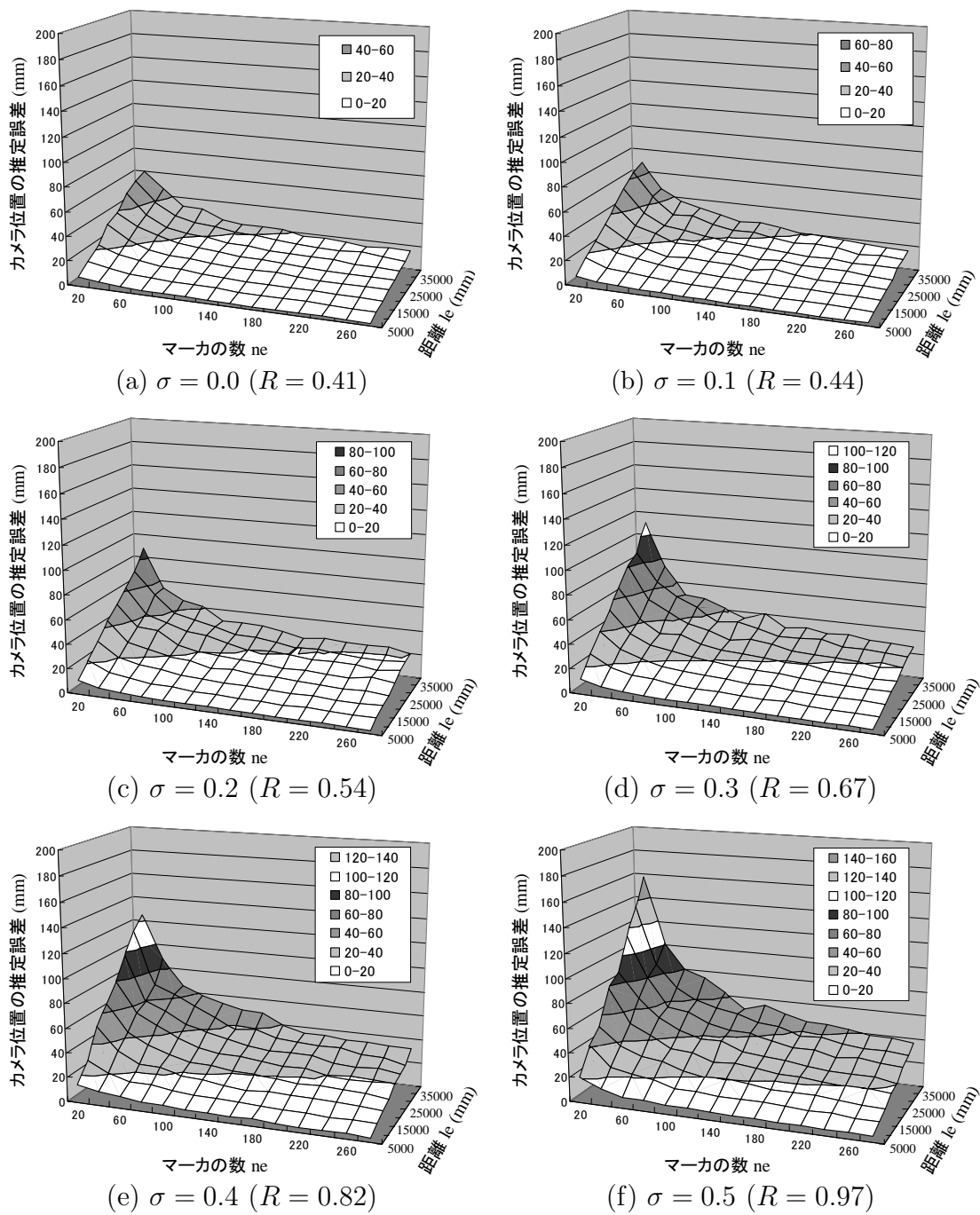


図 2.16 計算機シミュレーションによるフレーム毎のカメラ位置の推定誤差 (平面的なマーカ配置)

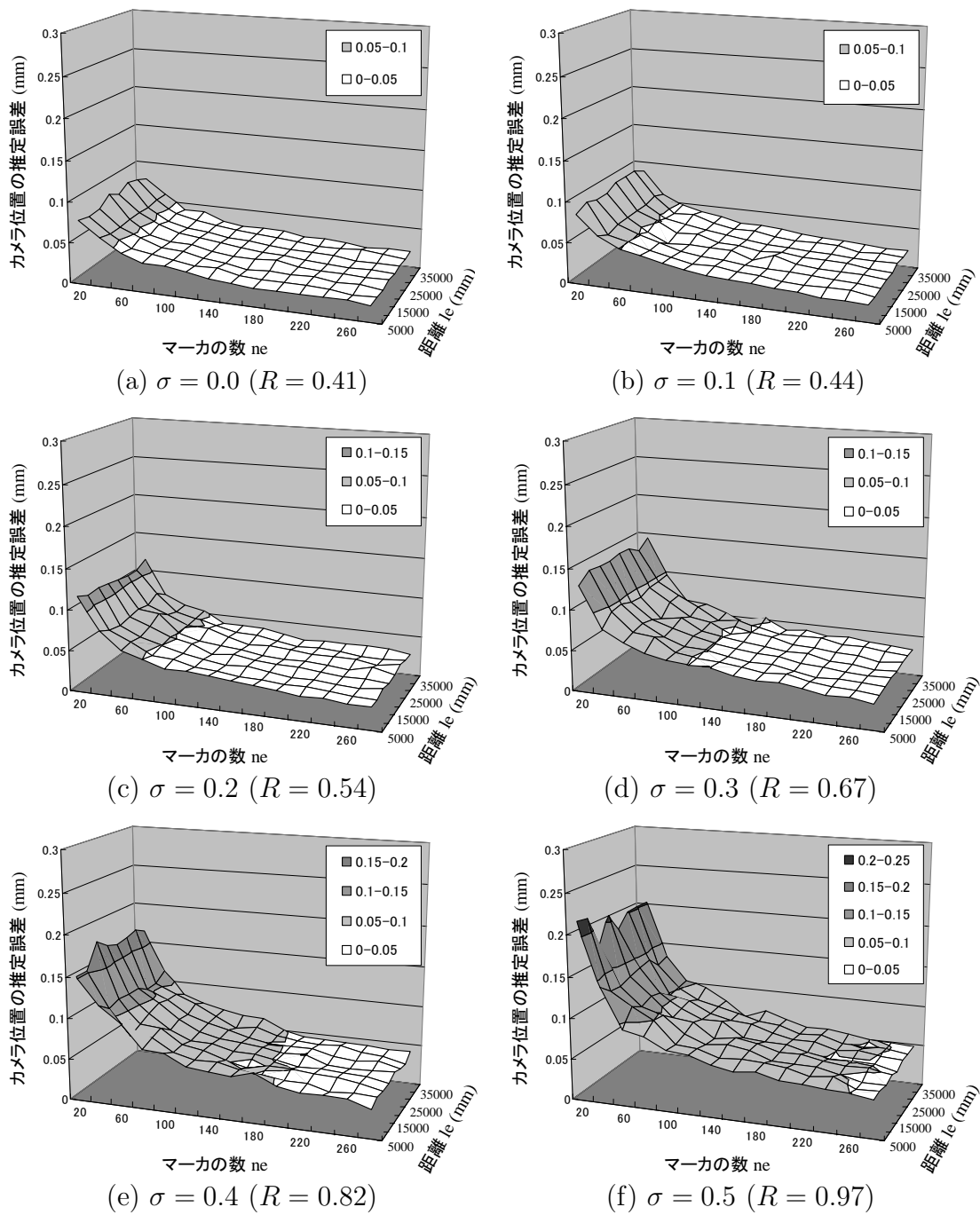
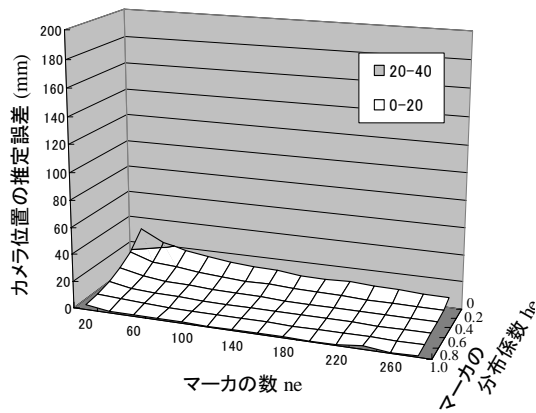
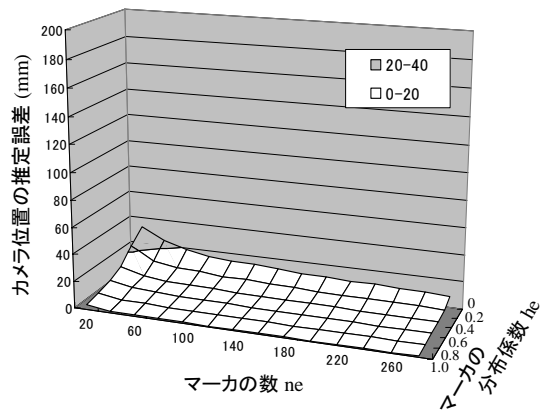


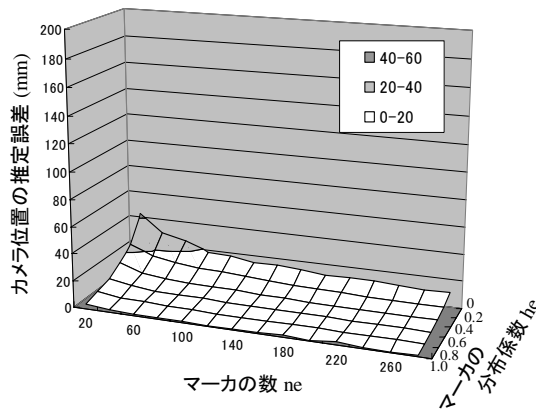
図 2.17 計算機シミュレーションによるフレーム毎のカメラ光軸方向の推定誤差 (平面的なマーカ配置)



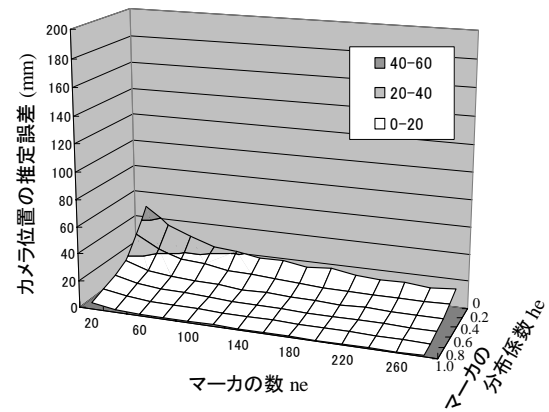
(a)  $\sigma = 0.0$  ( $R = 0.41$ )



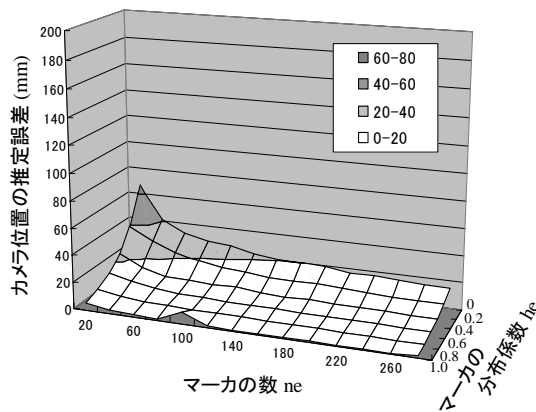
(b)  $\sigma = 0.1$  ( $R = 0.44$ )



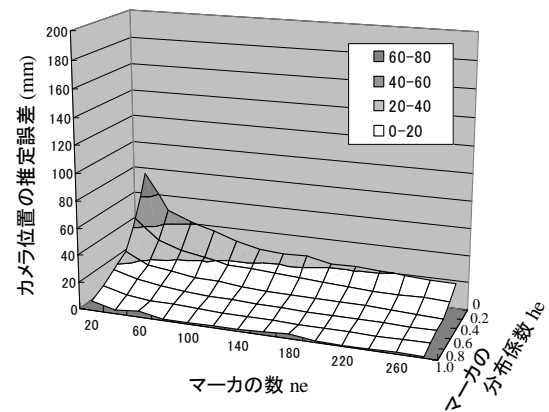
(c)  $\sigma = 0.2$  ( $R = 0.54$ )



(d)  $\sigma = 0.3$  ( $R = 0.67$ )

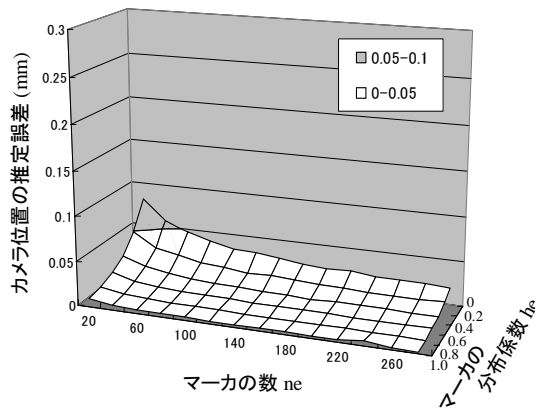


(e)  $\sigma = 0.4$  ( $R = 0.82$ )

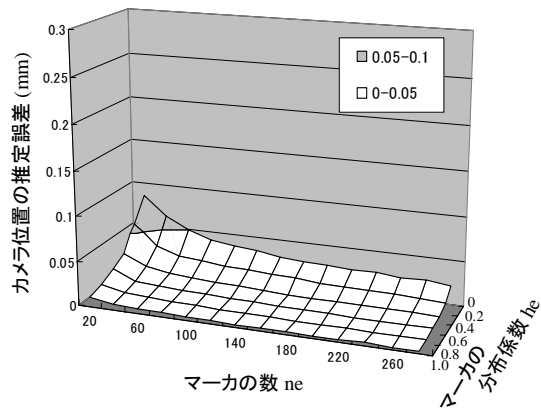


(f)  $\sigma = 0.5$  ( $R = 0.97$ )

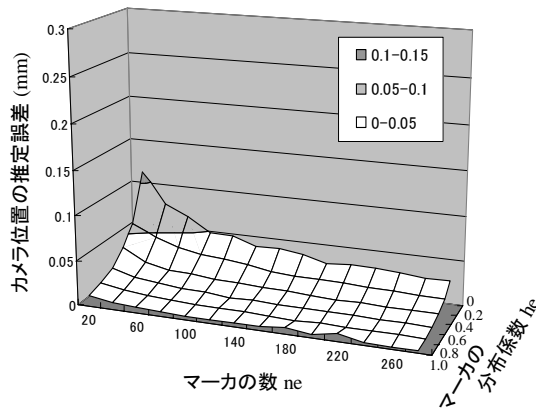
図 2.18 計算機シミュレーションによるフレーム毎のカメラ位置の推定誤差 (空間的なマーカ配置)



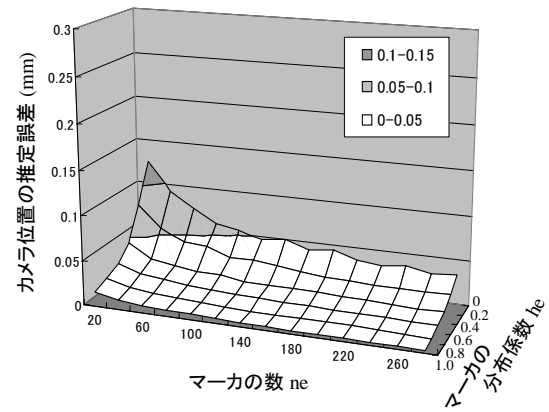
(a)  $\sigma = 0.0$  ( $R = 0.41$ )



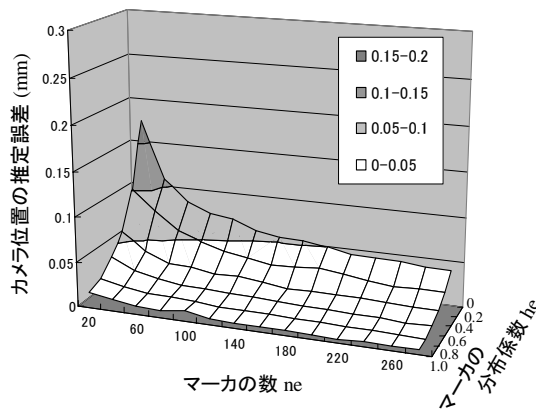
(b)  $\sigma = 0.1$  ( $R = 0.44$ )



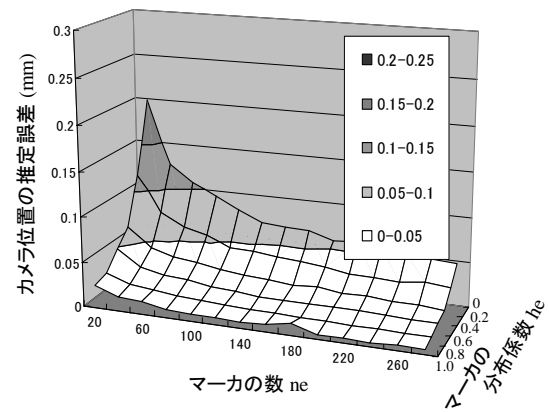
(c)  $\sigma = 0.2$  ( $R = 0.54$ )



(d)  $\sigma = 0.3$  ( $R = 0.67$ )



(e)  $\sigma = 0.4$  ( $R = 0.82$ )



(f)  $\sigma = 0.5$  ( $R = 0.97$ )

図 2.19 計算機シミュレーションによるフレーム毎のカメラ光軸方向の推定誤差 (空間的なマーカ配置)

## 2.8.2 カメラパラメータの累積的な誤差の評価

前節の実験 1, 2 により推定されたカメラパラメータの累積的な誤差の評価を行なう。屋外環境においてカメラパラメータの真値を得ることは困難であるため、以下では、まず全てのフレームの画像上で基準マーカを指定することで累積的な誤差を含まないカメラパラメータを推定し、実験 1, 2 で推定されたカメラパラメータとの比較を行なう。

### (1) 基準マーカのみによる累積誤差を含まないカメラパラメータの推定

全ての入力画像上で三次元位置が既知の基準マーカを手動で追跡することで、特徴点の誤追跡による累積的な推定誤差を含まないカメラパラメータを求める。ここでは基準マーカとして、2.7.1 項における実験 1 の画像上では各フレームにおいて平均 32 点、2.7.2 項における実験 2 では各フレームにおいて平均 15 点を指定し、2.5.1 項において述べたカメラパラメータの推定手法を用いることで、累積的な誤差を含まないカメラパラメータを推定した。

手動による特徴点の追跡においても、特徴点追跡の誤差および量子化誤差に起因するカメラパラメータの推定誤差が発生するため、前項で述べた計算機シミュレーションを行なうことで、この誤差の平均値を算出する。計算機シミュレーションに用いた基準マーカの条件を表 2.5 に示す。同表に示す条件は、実際に基準マーカのみを用いてカメラパラメータを推定した際に得られたデータである。ただし、マーカの分布範囲は、各フレームにおけるマーカの分布範囲の平均値として算出した。また、マーカの分布は各カメラの投影中心から光軸方向に対して、一定の範

表 2.5 累積的な誤差を含まないカメラパラメータの推定に用いた基準マーカの条件

	実験 1	実験 2
各フレームで用いたマーカ数の平均	32	15
再投影の誤差の平均値の平方 (画素)	0.63	0.85
マーカの分布範囲の平均 (m)	24 ~ 43	13 ~ 120

圏内で一様であると仮定した。表 2.6 に、計算機シミュレーションによって算出されたカメラパラメータの推定誤差の平均値を示す。この結果から、本実験で作成した累積的な誤差を含まないカメラパラメータには、実験 1 の入力に対して少なくとも平均 32mm、実験 2 の入力に対して平均 41mm 程度のカメラ位置の推定誤差が各フレームに含まれていることが分かる。同様に、カメラの光軸方向に対してもそれぞれ推定誤差が含まれているため、以下で述べる累積的な誤差評価では、これらの誤差を考慮する必要がある。

## (2) カメラパラメータの累積的な推定誤差の評価

実験 1, 2 により推定されたカメラパラメータと (1) で推定した累積的な誤差を含まないカメラパラメータを比較することで、累積的な推定誤差を評価する。実験 1 により推定されたカメラパラメータと、(1) で推定したカメラパラメータの各フレームにおけるカメラ位置の差を図 2.20 に示す。また各フレームにおける光軸ベクトルの角度の差をグラフで図 2.21 に示す。同図において、全体最適化前のカメラ位置の推定誤差は平均 240mm、光軸方向の推定誤差は平均 0.20°であり、全体最適化後にはそれぞれ、平均 100mm、平均 0.09°となった。このことから、2.6 節で述べた全体最適化によって、カメラパラメータ推定の累積誤差が軽減されていることが分かる。ただし、本実験で比較対象とした誤差の累積がないカメラパラメータには、表 2.6 に示した推定誤差が含まれているため、これら考慮すれば、全体最適化後の推定誤差はカメラ位置に対して平均 68 ~ 132mm、カメラ光軸方向に対して平均 0.04° ~ 0.14°の範囲であることが分かる。

図 2.22, 図 2.23 に、実験 2 で復元したカメラパラメータと基準マーカのみで復元されたカメラパラメータとの差を示す。実験 2 においては、全体最適化前のカメ

表 2.6 累積的な誤差を含まないカメラパラメータのフレーム毎の推定誤差の平均値

	実験 1	実験 2
カメラ位置 (mm)	32	41
カメラ光軸方向 (deg.)	0.05	0.03

ラ位置の推定誤差は平均 220mm，光軸方向の推定誤差は平均 0.22 °であり，全体最適化後にはそれぞれ，平均 90mm，平均 0.17 °となった．実験 1 と同様に，追跡誤差を考慮した全体最適化後の推定誤差はカメラ位置に対して平均 49 ~ 131mm，カメラ光軸方向に対して平均 0.14 ° ~ 0.20 °の範囲であることが分かる．

これらの結果から，一般的な屋外環境において 100m 程度の移動を伴う動画像からのカメラパラメータの復元では，平均 100mm 程度の位置の推定誤差が，カメラの姿勢に関しては，平均 0.1 °程度の推定誤差が発生することが分かる．この誤差は，表 2.6 に示したカメラパラメータの推定誤差に対して数倍程度大きいため，累積的な誤差が完全には削減されていないことが分かる．

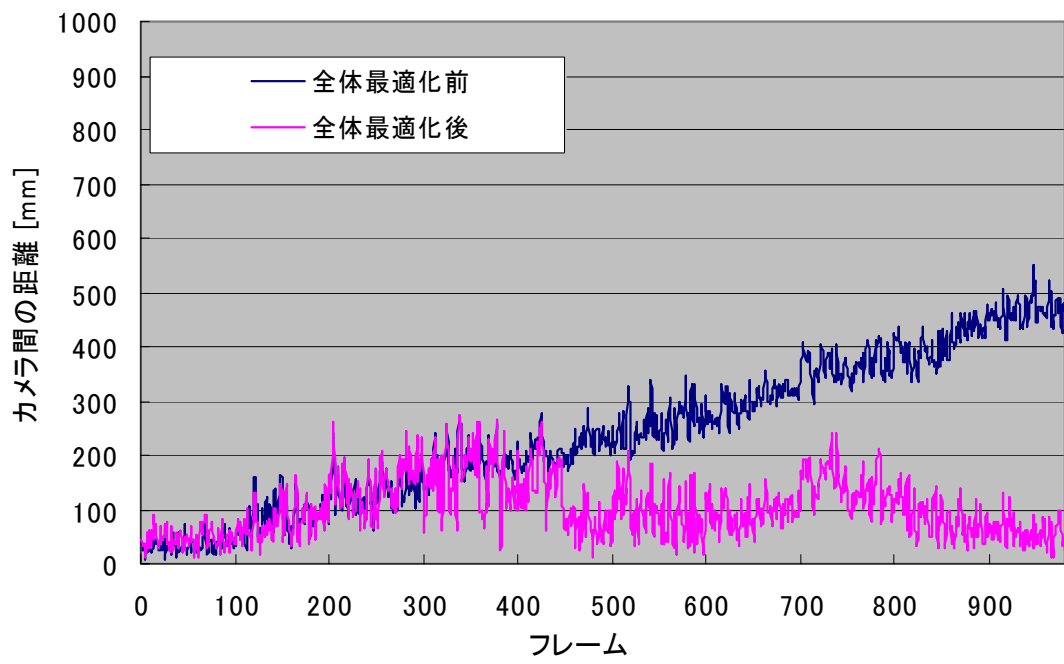


図 2.20 カメラ位置の推定誤差 (建物)

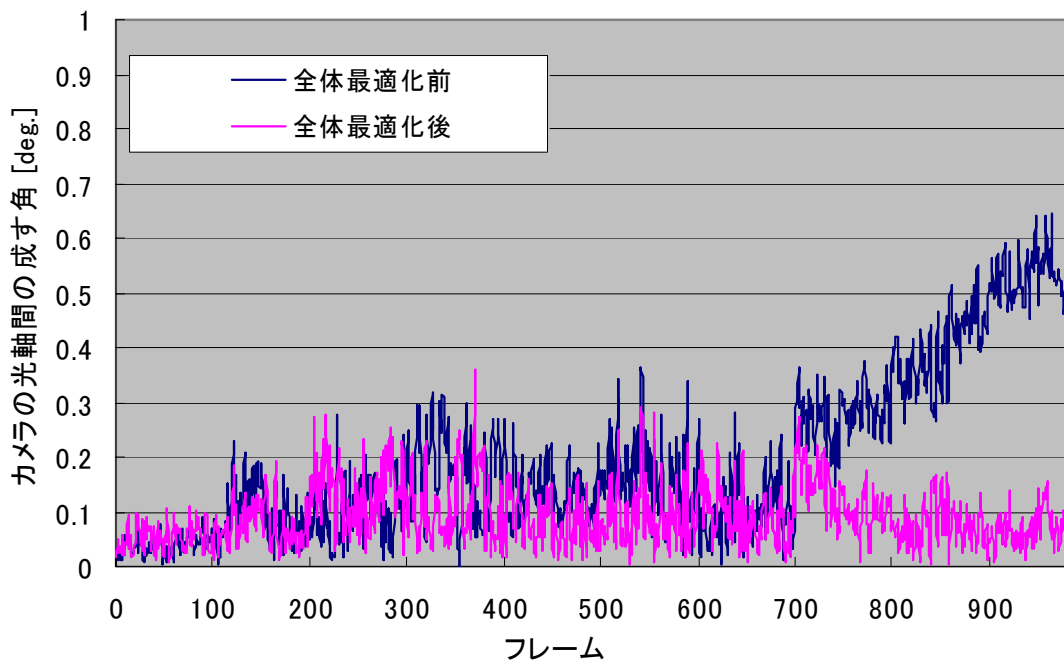


図 2.21 カメラ光軸の推定誤差 (建物)



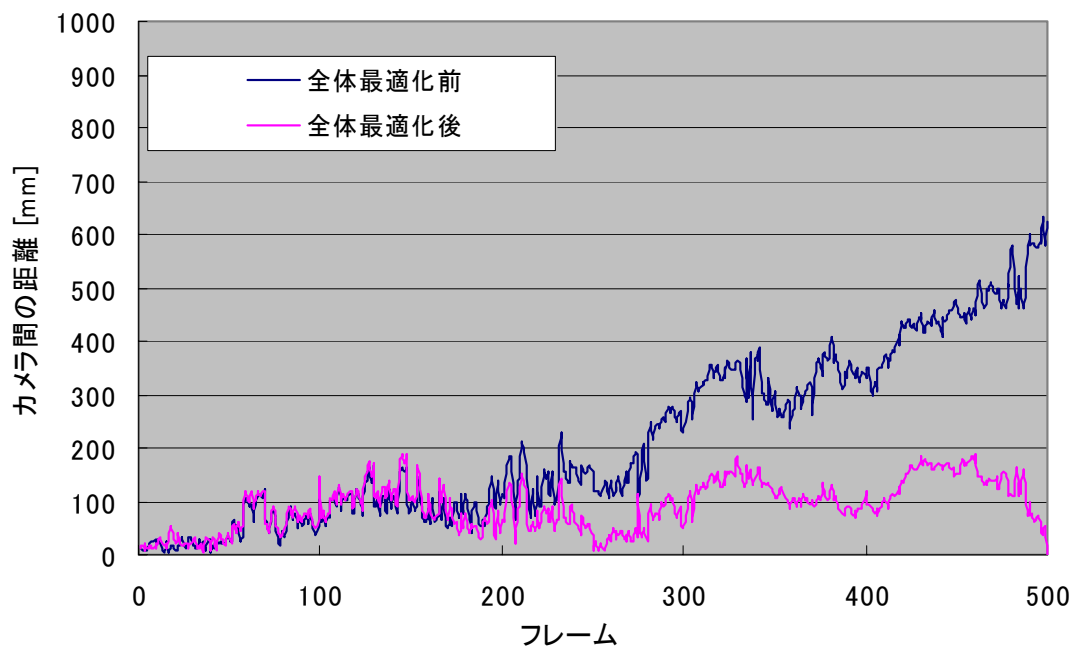


図 2.22 カメラ位置の推定誤差 (町並み)

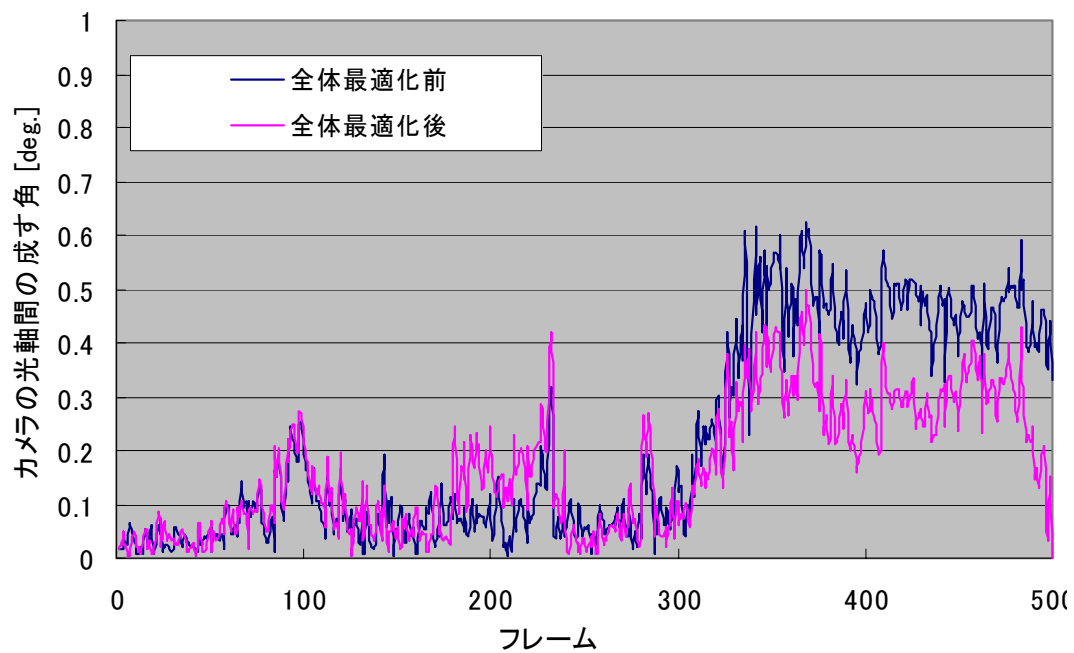


図 2.23 カメラ光軸の推定誤差 (町並み)

## 2.9 結言

本章では、まず基準マーカと自然特徴点を画像上で追跡することにより逐次的な三次元復元を行い、次に全体最適化によって累積的な推定誤差を最小化することで、精度良くカメラパラメータを推定する手法を提案した。実験により、提案手法では多くのオクルージョンを含む千枚程度の画像から成る動画像に対しても、累積誤差を抑えてカメラパラメータの復元を行うことが可能であることを確認した。

また、カメラパラメータの推定誤差に関する評価実験を行ない、以下の工夫を行なうことで、各フレームにおけるカメラパラメータの推定精度を向上させることが可能であることを確認した。

- (a) 特徴点を空間的に配置すること
- (b) 特徴点の追跡精度を向上させること
- (c) 使用する特徴点の数を増やすこと
- (d) 撮影対象までの距離を短くすること

カメラパラメータの累積的な誤差の評価実験では、一般的な屋外環境での 100m 程度の移動を伴う撮影において、本手法のカメラパラメータの推定誤差はカメラ位置に関して平均 100mm 程度、カメラ光軸方向に関して平均 0.1 °程度であることを確認した。この誤差は、マーカのみを用いて推定を行ったカメラパラメータの推定誤差に対して数倍大きいため、累積的な誤差が完全には削減されていないことが分かる。

このようなカメラパラメータの累積的な推定誤差は、各フレームにおける推定誤差が累積したものであると考えられる。そこで、今後、更なるカメラパラメータの推定精度の向上のためには、上述した工夫によるフレーム毎でのカメラパラメータの推定精度の向上や、全方位型のカメラを用いて追跡する特徴点の数や空間的な分布を大きくすることが有効であると考えられる。

## 3. 拡張マルチベースラインステレオ法を用いた奥行き推定に基づく三次元モデル化

### 3.1 序言

前章で述べたカメラパラメータの推定手法を用いることで、屋外環境を撮影した複数の動画像系列のカメラパラメータを、同一の座標系で精度良く推定することが可能となった。本章では、この手法によって推定されたカメラパラメータを用いて、拡張マルチベースラインステレオ法によって各フレームでの奥行きを密に推定し、共通座標系におけるボクセル空間への統合を行い三次元モデルを復元する手法について述べる。

以下では、まず提案する手法の基礎となるマルチベースラインステレオ法について詳述し、続いて複雑な屋外環境においても安定した奥行き画像を得るための、マルチベースラインステレオ法の拡張について述べる。次に、複数の動画像系列を用いて各フレームにおける奥行き画像の推定を行うことで、奥行き画像の復元精度を向上させる手法について詳述する。最後に、数百枚の奥行き画像を一つのボクセル空間に統合することで、三次元モデルを復元する手法について述べる。

### 3.2 マルチベースラインステレオ法の概要

本節では、提案する手法の基礎となる、マルチベースラインステレオ法 [OK93] について概要を述べる。マルチベースラインステレオ法は、任意に配置された校正済みの数台のカメラを用いて単一のシーンを撮影し、シーンの奥行き画像を推定する手法である。多数の画像を用いて一枚の奥行き画像の推定を行うために、二枚または三枚の画像を用いてステレオを行う手法に対して、安定した奥行き情報の推定が可能であるという特長がある。この手法では、複数台のカメラを用いていたが、以下では各カメラを動画像の各フレームで置き換えることにより、マルチベースラインステレオ法を動画像に適用する。

マルチベースラインステレオ法では、図 3.1 に示すように、第  $f$  フレームにおいて撮影される画素  $(x, y)$  の奥行き値  $z$  を、その周辺に配置された第  $j$  フレームから

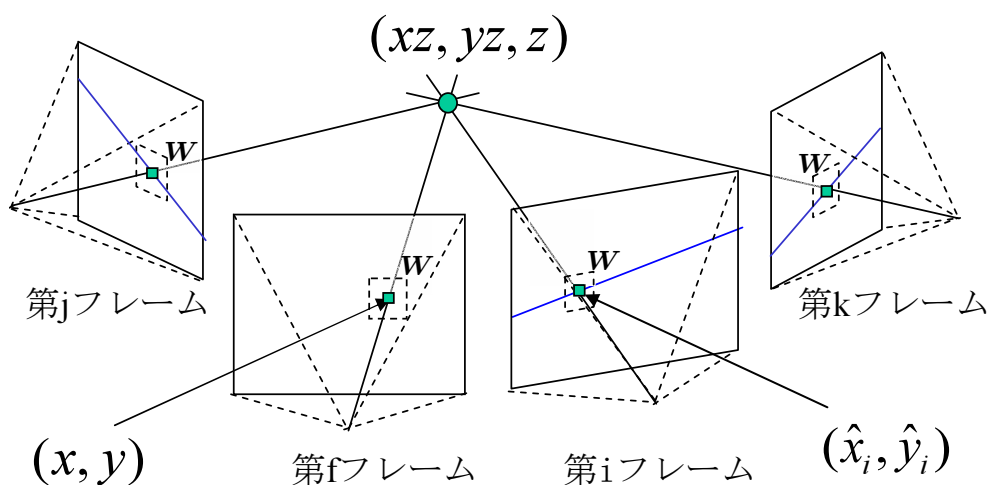


図 3.1 画素  $(x, y)$  の三次元位置と各画像上への投影直線

第  $k$  フレームの画像を用いて推定する ( $j \leq f \leq k$ ). 簡単のためにカメラの焦点距離を 1 とすれば, 第  $f$  フレームにおける画素  $(x, y)$  の三次元座標は  $(xz, yz, z)$  となり, 以下の式により, この三次元座標は第  $i$  カメラ ( $j \leq i \leq k$ ) の画像上  $(\hat{x}_i, \hat{y}_i)$  に投影される.

$$\begin{pmatrix} a\hat{x}_i \\ a\hat{y}_i \\ a \\ 1 \end{pmatrix} = \mathbf{M}_i \mathbf{M}_f^{-1} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix} \quad (38)$$

ただし,  $a$  は媒介変数,  $\mathbf{M}_f$  は第  $f$  フレームでの世界座標からカメラ座標への変換行列である. 図 3.1 に示すように,  $(\hat{x}_i, \hat{y}_i)$  は,  $(xz, yz, z)$  と第  $f$  フレームの投影中心を結ぶ直線を各探索画像面上に投影した直線上に拘束される. マルチベースラインステレオ法では, 第  $f$  フレームにおける画素  $(x, y)$  を中心とするウインドウ  $W$  と第  $i$  フレームにおける画素  $(\hat{x}_i, \hat{y}_i)$  を中心とするウインドウ  $W$  の輝度値  $I$  の差の二乗和 SSD (Sum of Squared Differences) を誤差の尺度として用いる.

$$SSD_{fi}(x, y, z) = \sum_{(u,v) \subseteq W} \{(I_f(x + u, y + v) - I_i(\hat{x}_i + u, \hat{y}_i + v))^2\} \quad (39)$$

SSD は、第  $f$  フレームと第  $i$  フレームの間のテンプレート誤差であるが、マルチベースラインステレオ法では、これらの和である SSSD(sum of SSD) を最小とするような奥行き値  $z$  を探索する.

$$SSSD_f(x, y, z) = \sum_{i=j}^k SSD_{fi}(x, y, z) \quad (40)$$

### 3.3 マルチベースラインステレオ法の拡張

前述したマルチベースラインステレオ法を用いて複雑な屋外環境を復元するには以下の問題がある.

#### 低信頼度領域での誤対応の問題

テクスチャの存在しない画像領域では、ウインドウ内での特徴が乏しいために、周辺領域への誤対応が発生し、奥行き推定の信頼度が低くなる.

#### オクルージョンとスケールの問題

撮影対象と各カメラの投影中心の間に遮蔽物が存在する場合、SSSD の算出に遮蔽物の影響が加わり、正しい奥行き値を算出することが困難となる. また、カメラ位置と対象物体の距離がフレームによって大きく異なるため、同一の物体であっても各カメラに投影される大きさが異なり、SSD が正しく算出されない.

#### 局所最小解と計算コストの問題

一定の奥行き推定分解能を保ったまま、多数の画像を用いて奥行き値の全探索を行うことは非常に計算コストがかかる問題である. また、勾配法的なアプローチを用いた場合には、多数の局所最小解が存在するため、誤った奥行き値が推定されてしまうという問題がある.

以下では、これらの問題を解決し、複雑な屋外環境においても安定した奥行き画像を得るための拡張について詳述する.

### 3.3.1 低信頼度領域の検出

二眼ステレオ法においては、一般的にエピポーラ線上において、エッジ上での対応点を優先的に決定することで信頼度の高いマッチングを行う手法が用いられている [Ard89, 山口 91, 島村 00]. 本手法では、これらの手法と同様のアプローチを用い、第  $f$  フレームにおいて画素  $(x, y)$  を中心とする窓  $W$  内において、輝度勾配の自乗和により算出される以下の評価値  $Grd$  が一定以下となる画素の奥行き値を内挿によって決定する.

$$Grd = \min \left( \sum_{(x,y) \in W} \left( \frac{\delta}{\delta x} I_f(x, y) \right)^2, \sum_{(x,y) \in W} \left( \frac{\delta}{\delta y} I_f(x, y) \right)^2 \right) \quad (41)$$

ただし、奥行き値の内挿は画像面に対して、水平方向、垂直方向の順で数回行ない、奥行き値が大きく異なる画素間での内層や、周辺に奥行き値が算出されていない画素に対する内挿は行なわない.

### 3.3.2 オクルージョンとスケールを考慮したマッチング尺度の算出

撮影対象と各カメラの投影中心の間に遮蔽物が存在する場合、SSSD の算出に遮蔽物の影響が加わり、正しい奥行き値を算出することが困難となる。また、カメラ位置と対象物体の距離が、フレームによって大きく異なるため、対象物体が異なる大きさで各カメラに投影される。

本研究では、前者に対して、ウインドウ間の誤差の中央値以下の和を SSSD として用いることで解決を図る。これにより、過半数のカメラ位置から三次元位置が可視であれば正しい奥行きを算出することが可能となる。また、後者に対して、スケールを考慮した SSD の算出を行うことで解決を図る。加えて、RGB の各要素の輝度値  $(I_R, I_G, I_B)$  を用いて SSD を算出することで、推定精度の向上を図る。以下に、これらを考慮して拡張したウインドウ間の誤差  $SSD'$  を示す。

$$\begin{aligned} SSD'_{fi}(x, y, z) = & \sum_{(u,v) \in W} \left\{ (I_{Rf}(x+u, y+v) - I_{Ri}(\hat{x}_i + su, \hat{y}_i + sv))^2 \right. \\ & + (I_{Gf}(x+u, y+v) - I_{Gj}(\hat{x}_i + su, \hat{y}_i + sv))^2 \\ & \left. + (I_{Bf}(x+u, y+v) - I_{Bj}(\hat{x}_i + su, \hat{y}_i + sv))^2 \right\} \quad (42) \end{aligned}$$

ここで,  $(\hat{x}_i, \hat{y}_i)$  は, 第  $f$  フレームにおけるカメラ座標系での三次元座標  $(x, y, z)$  を第  $i$  フレームへ投影した画像座標,  $s$  はスケール係数であり, 以下の式により算出する.

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathbf{M}_i \mathbf{M}_f^{-1} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix} \quad (43)$$

$$\hat{x}_i = X/Z \quad (44)$$

$$\hat{y}_i = Y/Z \quad (45)$$

$$s = z/Z \quad (46)$$

また,  $SSD'$  のメディアン値  $M$  を用いて  $SSSD'$  を以下のように定義する.

$$SSSD'_f(x, y, z) = \sum_{i=j}^k \begin{cases} SSD'_{fi}(x, y, z); & SSD'_{fi}(x, y, z) \leq M \\ 0; & \text{otherwise} \end{cases} \quad (47)$$

ただし,

$$M = \text{median}(SSD'_{fj}(x, y, z), \dots, SSD'_{fk}(x, y, z)) \quad (48)$$

このような  $SSSD'_f$  を最小化することで, 奥行き探索に用いる画像の半数以上で可視の画素  $(x, y)$  であれば, 奥行き値  $z$  を正しく推定することができる.

### 3.3.3 多重スケール法の利用

前述した  $SSSD'$  を用いた奥行き値  $z$  の探索において, 一定の奥行き推定分解能を保ったまま, 多数の画像を用いて奥行き値の全探索を行うことは非常に計算コストがかかる問題である. また, 勾配法的なアプローチを用いた場合には, 多数の局所最小解が存在するため, 誤った奥行き値が推定されてしまうという問題がある. 本論文では, これらの問題を解決するために, 多重スケール法 [横矢 93] を利用し, 入力画像に対する複数の解像度を段階的に用いて奥行き値  $z$  の探索を行う.

多重スケール法では, 図 3.2 に示すような, 各解像度の画像を層とする多解像度の画像ピラミッドを用いる. 本手法では, 各層の解像度が下層の解像度の  $1/2$  と

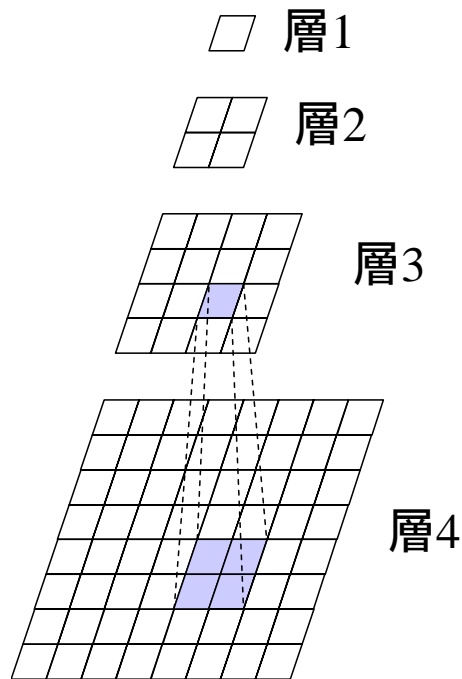


図 3.2 画像ピラミッドの例

なるように画像ピラミッドを構成し、各層における 1 画素は、下層の 4 画素と直接対応している。このような多重解像度のピラミッドを用いて、まず最も粗い解像度の画像（ピラミッドの上位層）に対して、あらかじめ設定した探索範囲内で奥行き値  $z$  を算出する。次の層では、上位層の対応する画素の周辺で求めた奥行き値の最大値と最小値を算出し、この奥行き値の最大値と最小値の間を含む限定された範囲内のみで探索を行なう。これを順次最下層まで繰り返して奥行き画像を推定することにより、 $z$  の探索における局所最小解の問題を回避する。

各層における奥行き値  $z$  の探索ステップとしては、奥行きの探索範囲が最も大きく投影される探索画像上でのサブピクセル単位での視差を用いる。これによって、ピラミッドの各層での奥行き値の探索ステップが画像の解像度に応じたものとなる。また、より解像度の高いピラミッドの下層では、上位層によって探索範囲が限定されるために、多重スケール法を用いない場合に比べて奥行きの探索にかかる計算コストを大幅に削減することができる。



### 3.4 複数の動画像系列を用いたマルチベースラインステレオ

町並みのような複雑な屋外環境においては、遮蔽物が多数存在するために、復元対象の一部が遮蔽物によって隠蔽されることが頻繁に発生する。このような遮蔽物によって隠蔽されるモデルの一部においては、推定に用いるベースライン距離が短くなるために奥行き推定の精度が低下し、復元される三次元モデルに欠落が生じる。そこで本研究では、現実環境をカメラの移動経路・向きを変えて複数回撮影した複数の動画像系列を用いて、前述した拡張マルチベースラインステレオ法による奥行き推定を行うことで、遮蔽物によるモデルの隠蔽を減少させ、モデルの推定精度を向上させる。

ただし、複数の動画像系列に対してマルチベースラインステレオ法に適用するためには、以下の2つの問題が発生するため、本節では、これらの問題を解決する手法について詳述する。

#### 動画像間での日照条件の変化の問題

一般的に、時間を置いて撮影した複数の動画像間では、日照条件の変化によって撮影対象の明るさが変化する。拡張マルチベースラインステレオ法では、輝度値の差の自乗和を誤差尺度として奥行き情報の推定を行うために、明るさが変化する複数の動画像系列を用いて正しい奥行きを算出することは困難である。

#### 奥行き探索に用いる画像選択の問題

拡張マルチベースラインステレオ法において、対象物体の正しい奥行き値を推定するためには、奥行き探索に用いる半数以上の画像上で対象物体が可視である必要がある。奥行きを正しく算出するためには、複数の動画像系列に対してこの条件を満たすような画像を自動で選択する必要がある。

#### 3.4.1 動画像間での日照条件の変化による影響の削減

拡張マルチベースラインステレオ法では、式(42)に示した画像上の輝度値の差を尺度として奥行き推定を行うために、日照条件の変化などによって複数の動画像間で対象物体の色が変化する場合には、これらを用いた精度の高い奥行き

推定を行うことは難しい。そこで、撮影対象を拡散反射成分のみから成ると仮定し、複数の動画像間で同一の物体の色を数点サンプリングすることによって、日照条件の変化による影響を入力画像から削減する。

本手法では、撮影対象の拡散反射成分のみを考慮することによって、照明条件の変化による画像間での同一の物体の色の変化を以下のような線形変換として近似する。

$$(I'_R, I'_G, I'_B)^T = (s_r, s_g, s_b)(I_R, I_G, I_B)^T + (o_r, o_g, o_b) \quad (49)$$

ただし、 $(I'_R, I'_G, I'_B)^T$  および  $(I_R, I_G, I_B)^T$  は照明条件の変化の前後での画素の色を、 $(s_r, s_g, s_b)$  および  $(o_r, o_g, o_b)$  は変換係数を表す。この変換係数は、複数の動画像間で同一の物体の色を数点サンプリングすることによって、最小自乗法によって算出する。この変換係数を用いて各動画像系列の画像を、奥行き探索を行う動画像系列のものに変換することで、日照条件が異なる動画像間においても正しい奥行き値の推定が可能となる。

### 3.4.2 奥行き探索に用いる画像の自動選択

3.3.2項で述べたように、拡張マルチベースラインステレオ法において対象物体の奥行き値を正しく推定するためには、奥行き探索に用いる半数以上の画像上で対象物体が可視である必要がある。本手法においては、多重スケール法によって限定される探索範囲を用いることによって、カメラの移動経路・向きが異なる複数の動画像系列において、これらの条件を満たすような奥行き画像の推定に用いる画像を自動で選択する。

本手法では、図 3.3 に示すように、ピラミッドの各階層において限定された探索範囲を各カメラ位置の画像上に投影し、探索範囲が全て画像内に収まるカメラの画像を選択し、画素の奥行き推定に用いる。ただし、同図中の曲線はカメラの移動の軌跡を、三角はカメラの姿勢を表す。たとえば、同図のカメラ位置  $c$  上の画素  $(x, y)$  における奥行きは、カメラ位置  $d, f, g, h$  の画像を用いて推定される。これにより、探索範囲が可視でないカメラ位置の画像を排除し、複数の動画像系列を用いた場合においても、探索に用いる画像を自動的に選択することができる。

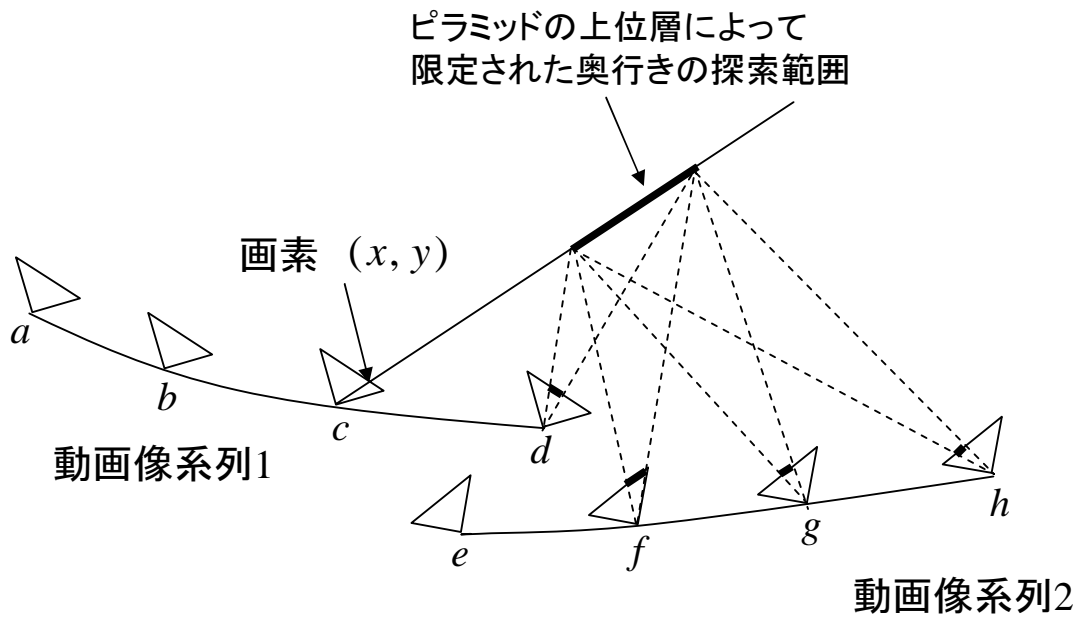


図 3.3 奥行き探索に用いる画像の選択

### 3.5 ボクセル空間におけるモデルの復元

これまで述べた手法を用いることで、複数の動画像系列における全てのフレームの奥行き画像を推定することが可能である。本節では、これらの奥行き画像を用いてボクセル空間への投票を行なうことで、ボクセルモデルとして三次元モデルを復元する手法について述べる。

まず、ボクセル空間について述べる。ボクセル空間とは、三次元空間を立方体(ボクセル)を単位とする格子で分割することで表現される空間であり、ボクセル空間を用いた三次元表現は、医療などの分野で内部構造の可視化の手法として研究されてきた。本手法においては、あらかじめ複数の動画像系列においてカメラパラメータの世界座標系が統一されているため、この世界座標系を用いて復元対象が全て含まれる領域を指定し、復元に用いるボクセル空間を設定する。

次に、動画像の各フレームにおける奥行き情報を用いて、設定したボクセル空間への投票を行なう。ここでは、各ボクセルは  $A$ ,  $B$  二つの投票箱を持つものとする。図 3.4 に示すように、まず、奥行き値が推定された画素を、その奥行き値  $z$  を用いてボクセル空間に逆投影し、対応するボクセルの投票箱  $A$  に投票する。同時に、カメラの投影中心と投票箱  $A$  に投票されたボクセルの間に存在するボクセルに対して投票箱  $B$  に投票を行う。このような投票において、投票箱  $A$  に投票されたボクセルは実際に物体として存在している領域である可能性が高く、逆に投票箱  $B$  に投票されたボクセルは物体が存在していない領域である可能性が高い。そこで、このような投票を全てのフレームの全ての画素に対して行い、評価値(投票箱  $A$  の投票値)/(投票箱  $B$  の投票値)が設定した閾値を超えるボクセルを、物体が存在する領域とすることで三次元モデルを復元する。また、各ボクセルの色は、投票箱  $A$  に投票した画素の色の平均を用いることで決定する。このような投票を行なうことで、複数の動画像上の奥行き画像を単一のボクセル空間に統合し、撮影対象をテクスチャ付きのボクセルモデルとして復元することが可能となる。

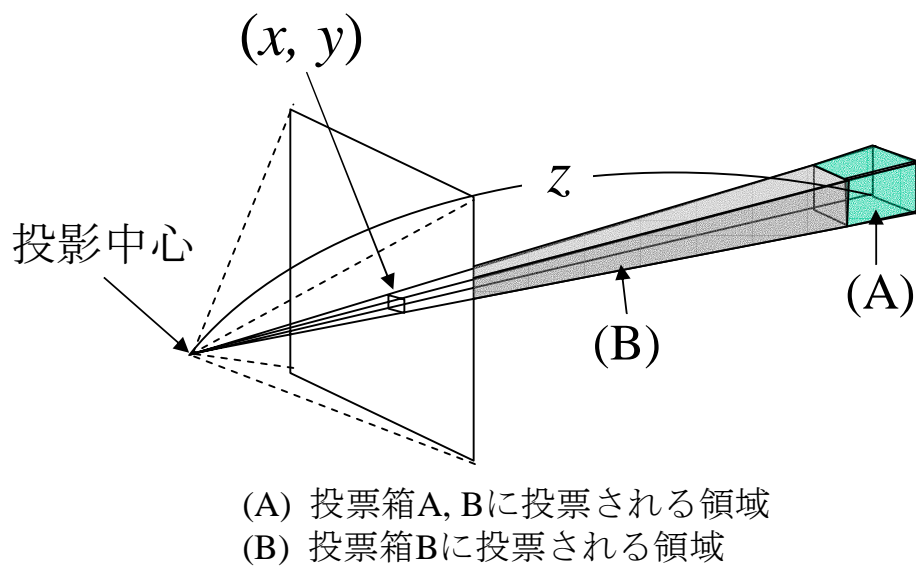


図 3.4 画素  $(x, y)$  の奥行き推定値  $z$  によるボクセル空間への投票

### 3.6 実験と考察

提案手法の実環境での有効性を確認するため、前章の実験と同様に、奈良・朱雀門と大阪・住道の町並みをビデオカメラで撮影し、複数の動画像を入力とする屋外環境の復元実験を行った。表 3.1 に、本実験で使用した実験条件を示す。また、カメラパラメータの推定における実験条件は 2 章で示したものと同一である。

表 3.1 奥行き画像の復元とボクセル空間への統合における実験条件

奥行き画像の復元のパラメータ

	建物を対象 とした実験	町並みを対象 とした実験
奥行き探索範囲 (m)	10 ~ 40	5 ~ 30
多重スケール法における 画像ピラミッドの層数	4	4
多重スケール法における 最上層の解像度 (画素)	45 × 30	45 × 30
多重スケール法における 最下層の解像度 (画素)	360 × 240	360 × 240

ボクセル空間の解像度とサイズ

	建物を対象 とした実験	町並みを対象 とした実験
ボクセル空間の解像度	450 × 260 × 240	1000 × 180 × 110
単位ボクセルのサイズ (cm)	10×10×10	10×10×10

### 3.6.1 建物を対象とした三次元モデル化 (実験 1)

本実験では、ワイドレンズ (Sony VCL-HG0758) を取り付けたビデオカメラ (Sony DSR-DP-150) を用いて、建物の正面・背面を二つのシーケンスに分けて撮影し、入力動画像系列として用いた。これらの動画像系列は、図 3.5 を含む (a) 建物正面の画像 747 枚と (b) 建物背面の画像 982 枚 (720×480 画素, プログレッシブ撮影) から成る。また、2つの動画像系列のカメラパラメータは、2章で述べた手法を用いることで推定した。ただし、あらかじめカメラパラメータの復元に用いる基準マーカの三次元位置を、トータルステーション (Leica TCR1105XR) を用いて単一の座標系で計測しておくことで、二つの動画像系列のカメラパラメータの座標系を統一した。図 3.6 に、同一の座標系で推定された二つのカメラパラメータを示す。

以下では、この入力動画像系列と推定されたカメラパラメータを用い、まず予備実験として奥行き推定における SSD 算出時のウィンドウサイズについての検討を行ない、次に、奥行き画像の推定結果をボクセル空間へ統合することで三次元モデルを復元する。ただし、本実験では、二つの動画像系列間において共通して撮影されている部分が少ないため、日照条件の影響は考慮しない。また、奥行き画像の推定に用いる画像は、3.4.2 項で述べた手法によって自動で選択する。

#### (1) SSD 算出のウィンドウサイズに関する予備実験

上述したカメラパラメータを推定済みの動画像系列を用いて、奥行き画像を算出するために適したウィンドウサイズについて検討する。ここではまず、図 3.7 に示す動画像中の単一のフレームに対して、(a) 1×1, (b) 3×3, (c) 5×5 の異なる 3つのウィンドウサイズを用いて SSD を算出し、奥行き画像の推定結果を比較した。図 3.8(a), (b), (c) に、それぞれのウィンドウサイズを用いて推定した奥行き画像を示す。同図に示した奥行き画像は、推定された奥行き値を輝度値に変換することで奥行き情報を可視化したものである。本実験において、PC (Pentium3 1GHz, Memory 1GB) を用いて奥行き画像の算出に要した時間は、それぞれ、460 秒 (a), 950 秒 (b), 1700 秒 (c) であった。

図 3.7 および図 3.8(a), (b), (c) から、ウィンドウサイズの大きい (b), (c) では、おおむね正しい奥行き情報が推定されていることを確認できるが、復元対象の細

部の凹凸が再現されず、つぶれてしまっていることが分かる。ウィンドウサイズを最小に設定した (a) では、全体的に誤推定が多く、屋根の一部や建物の柱の間などに明らかに誤った奥行き推定結果が見られる。しかし、建物の欄干や画像下部の柵の部分のように形状が細かく、かつ不連続となるような部分において、奥行き値がつぶれることなく推定されている。

ウィンドウサイズを最小に設定した (a) で見られる奥行き推定の誤りは、多重スケール法における画像ピラミッドの上位層において、奥行き値の探索範囲内に多くの類似色が存在するために誤推定が発生し、誤った奥行き値の推定結果が下層に伝播するために発生していると考えられる。そこで、画像ピラミッドの最下層では $1 \times 1$ のウィンドウを、その他の層においては $3 \times 3$ のウィンドウを用いることで、奥行き画像の推定精度の向上を試みた。この異なるウィンドウサイズを用いた奥行き推定の結果を図 3.8(d) に示す。同図 (d) に示した推定結果から、(a) の推定結果に見られた推定の誤りは大幅に減少しており、加えて建物の欄干のような細かな形状の凹凸に対しても奥行き値が奥行き値がつぶれることなく推定されている。また、同図 (d) の奥行き画像の算出に要した時間は 380 秒であり、(a) ~ (d) の中で最も高速であった。これは、画像ピラミッドの上位層において誤った奥行き値の推定結果が減少することで、下層での奥行き探索範囲が小さい範囲に限定されたためであると考えられる。

以上の結果から、建物を対象とした実験においては、(a)、(b)、(c) のように、画像ピラミッドの各層で同一のウィンドウを用いる場合に比べて、上位層では大きなウィンドウ、下位層では小さなウィンドウといった、可変のウィンドウを用いて SSD を算出する方が、推定精度の向上および計算コストの削減に対して有効であると考えられる。

## (2) 三次元モデルの復元実験

各フレームにおいて密に奥行きを推定し、ボクセル空間への統合を行なうことで、三次元モデルを復元する。(1) の予備実験での結果から、本実験では、多重スケール法における奥行き値の探索に、最下層で $1 \times 1$ 、その他の層で $3 \times 3$ のウィンドウを用いる。まず、3.3 節に述べた手法により、二つの入力動画像上の全てのフレームで奥行き画像を推定した。ただし、3.4.2 項で述べた探索に用いる画像の自



動選択で、最上位層での画像の選択枚数が少数となるフレームに対しては、ベースライン距離が短く、探索範囲の限定が困難となるため、奥行き画像の推定を行わなかった。図 3.9 は推定された画素の画像の一部である。同図から、予備実験と同様に安定して奥行きが推定されていることが確認できるが、建物の正面に対して平行に移動して撮影されたフレームの周辺において、屋根などの縦方向のエッジを含まない部分に誤った奥行き値が見らる。また、本実験では、24 台の PC から成る PC クラスタ (Pentium4 1.8GHz, 1GB Memory) を用いて並列で奥行き画像を推定し、約 1700 枚の奥行き画像を算出するのに要した時間は 280 分であった。

次に、現実空間において一辺が 10cm に対応するボクセルで構成されたボクセル空間 (解像度 :  $450 \times 260 \times 240$ ) に、3.5 節に述べた統合手法を用いて、図 3.9 を含む奥行き画像約 1700 枚を統合した。統合して得られたテクスチャ付き三次元モデルを図 3.10 に示す。同図から、建物の柱のようにオクルージョンがおこる部分においてもおおむね正しくモデルが復元されていることが分かる。しかし、建物の内側の壁面のように、動画像上での隠蔽が多い部分において復元のエラーが見られる。また、建物側面や上方部分は奥行き値の推定枚数が少ないために復元されなかった。



第1フレーム



第250フレーム



第500フレーム



第747フレーム

(a) 建物正面



第1フレーム



第325フレーム



第650フレーム



第982フレーム

(b) 建物背面

図 3.5 入力画像 (建物)

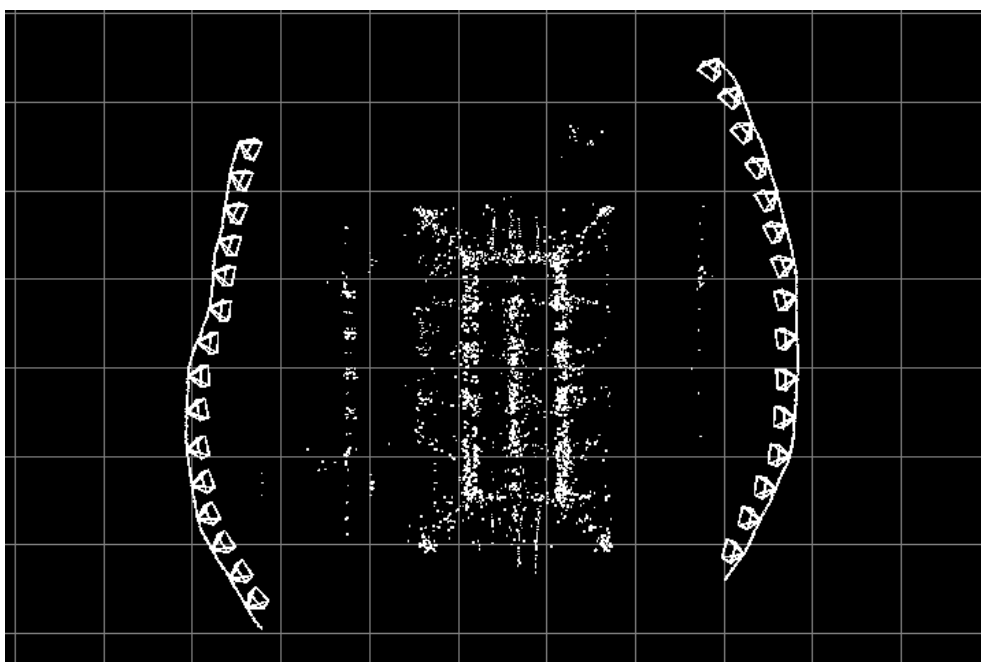
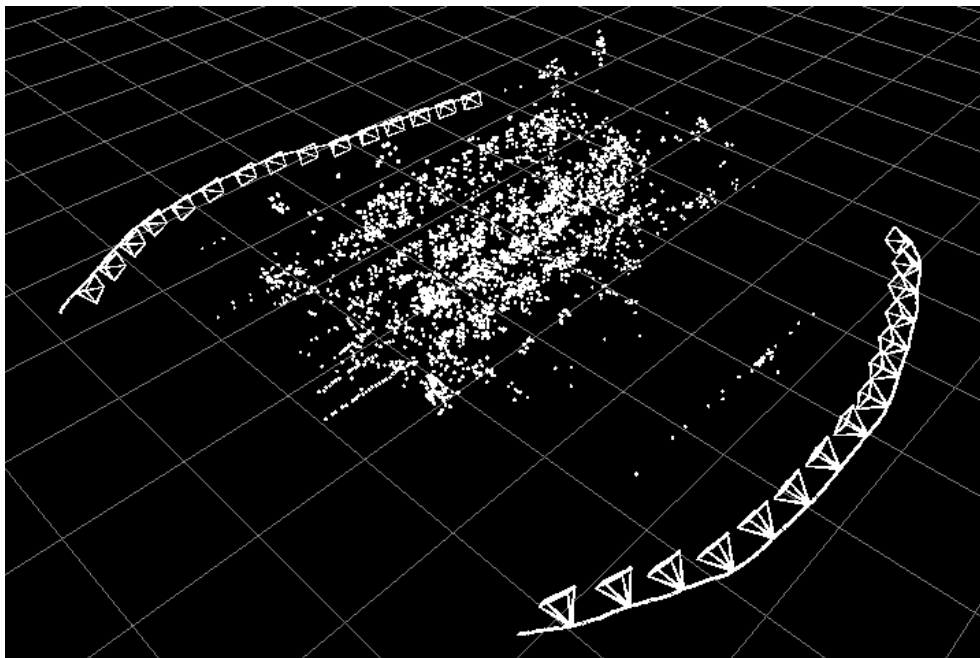


図 3.6 復元されたカメラパラメータと自然特徴点の三次元位置 (建物)



図 3.7 SSD 算出のウィンドウサイズによる奥行き推定結果の比較に用いた入力フレームの画像 (建物)

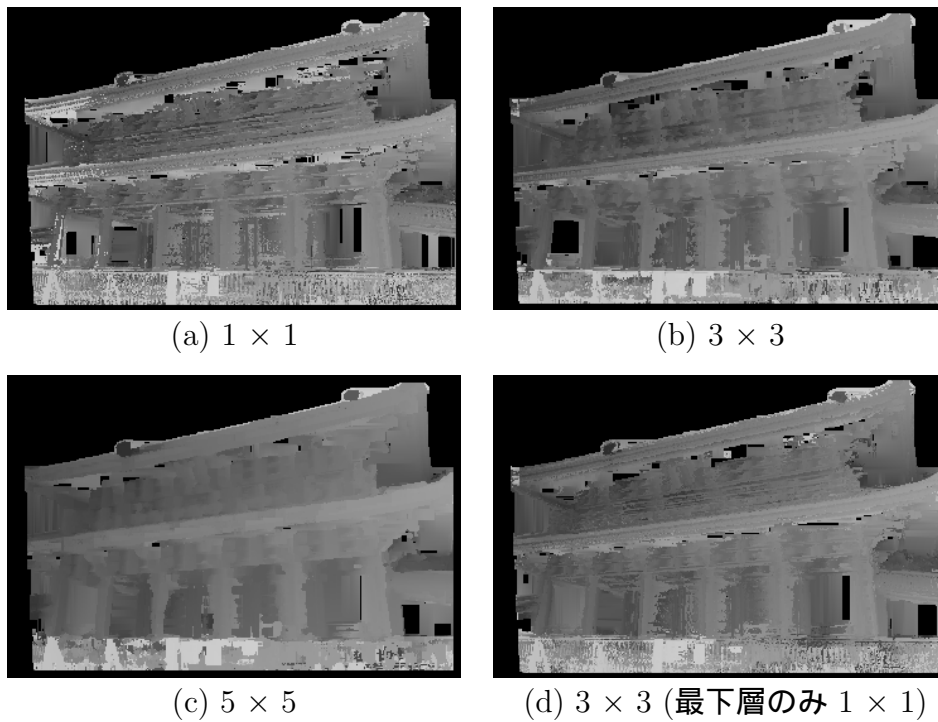
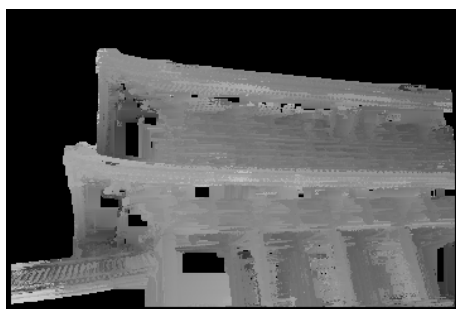
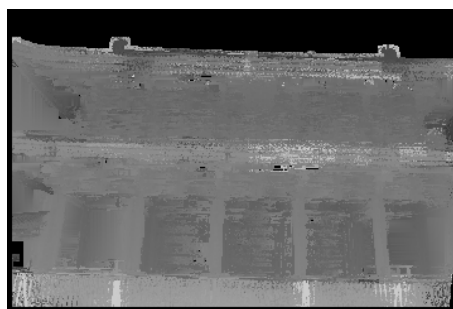


図 3.8 SSD 算出のウィンドウサイズによる奥行き画像の比較 (建物)



第 100 フレーム



第 280 フレーム



第 460 フレーム



第 644 フレーム

(a) 建物正面



第 100 フレーム



第 360 フレーム



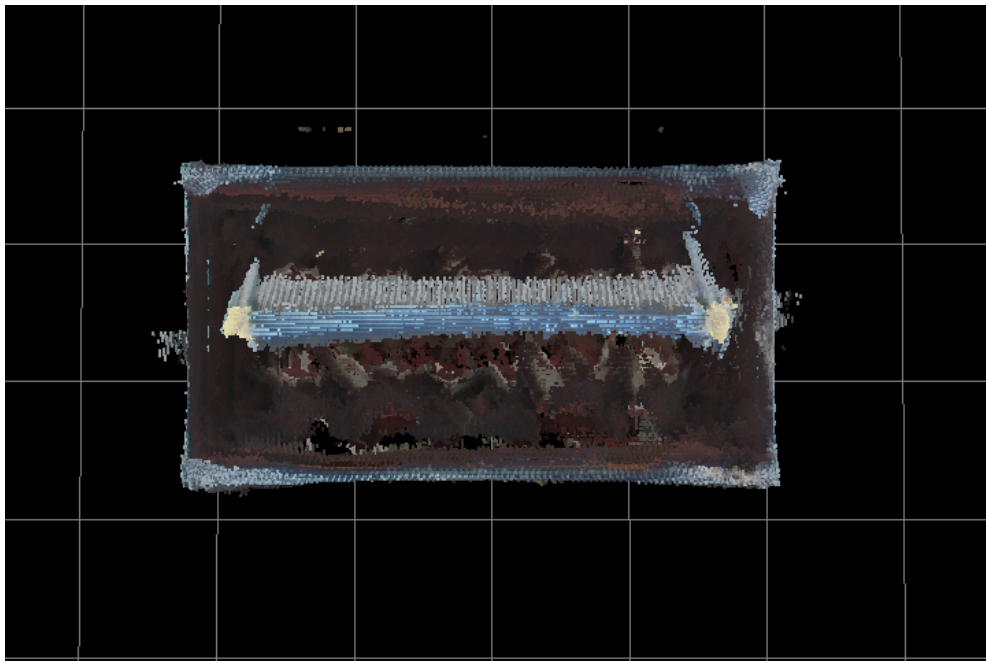
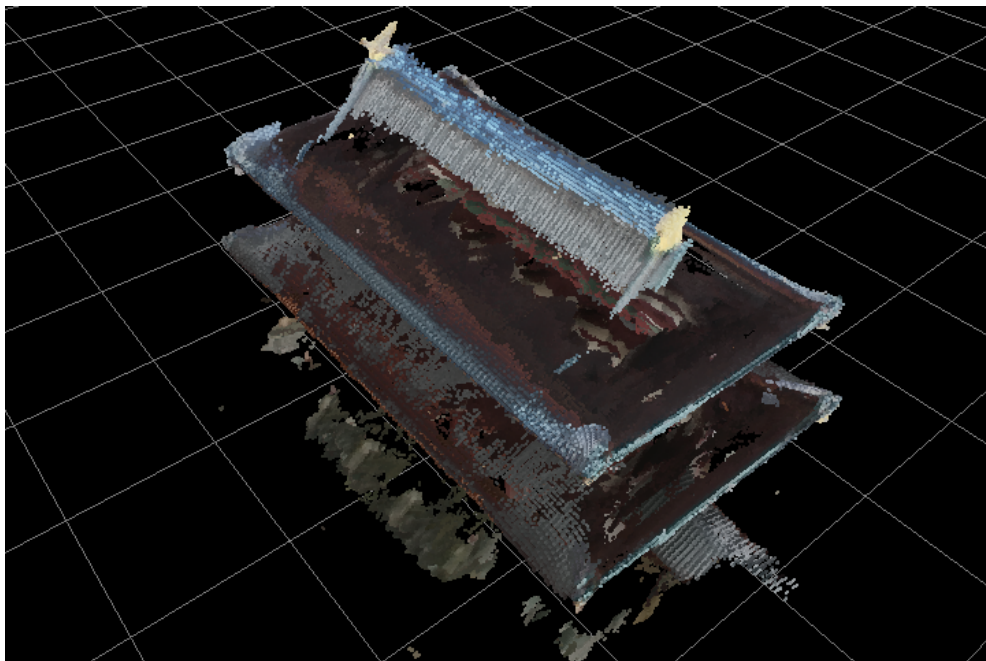
第 620 フレーム



第 880 フレーム

(b) 建物背面

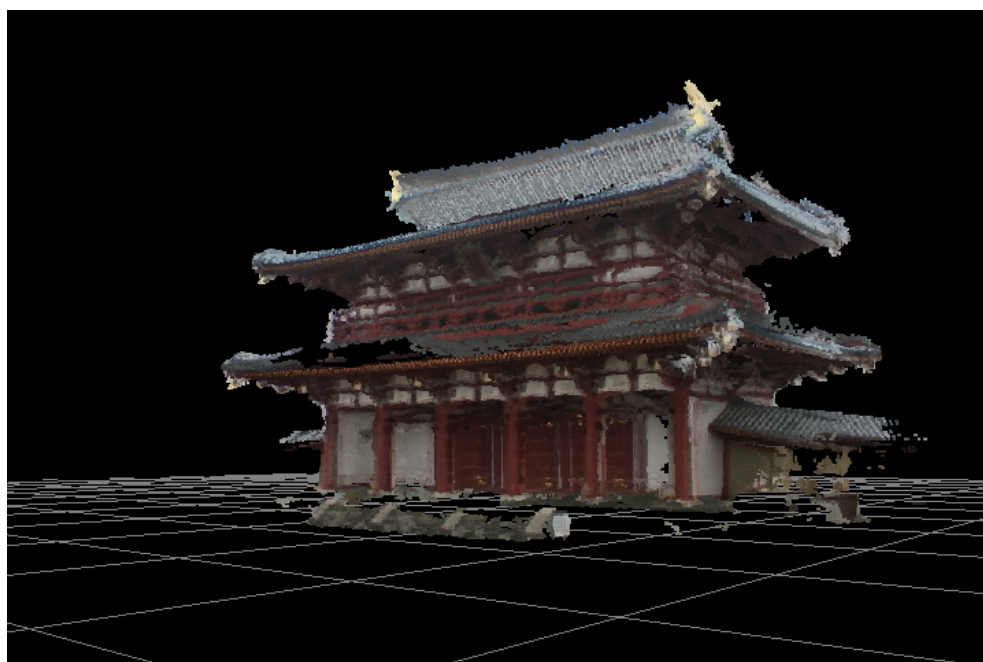
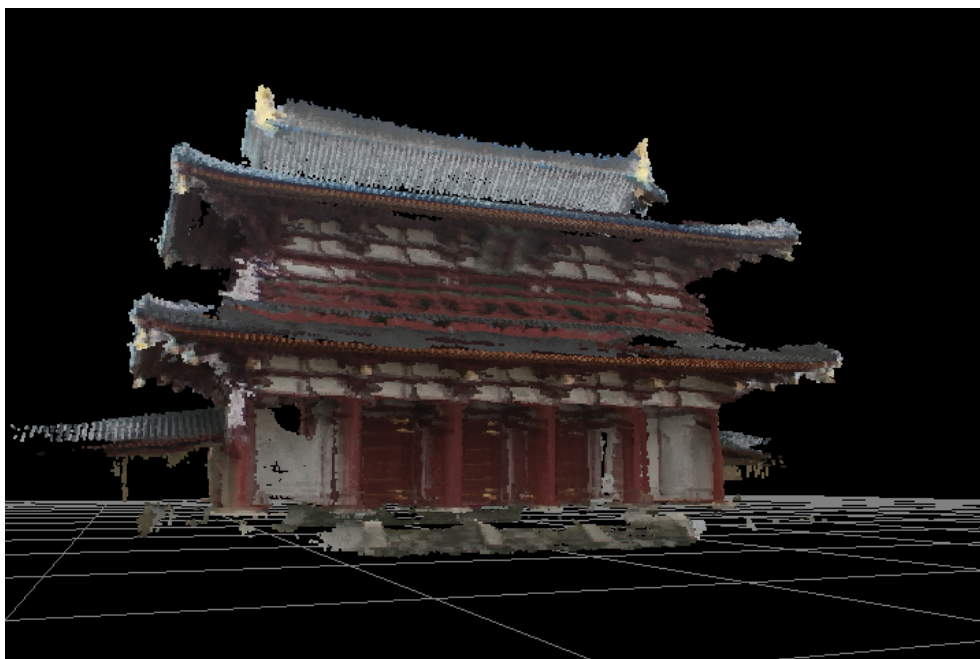
図 3.9 奥行き画像の推定結果 (建物)



(a) 建物上方

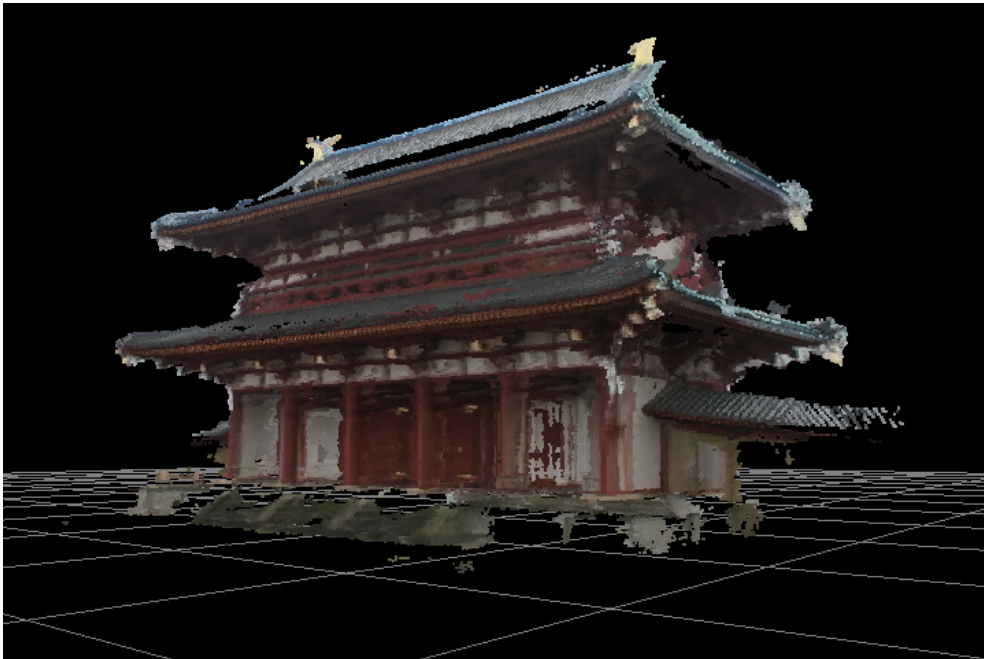
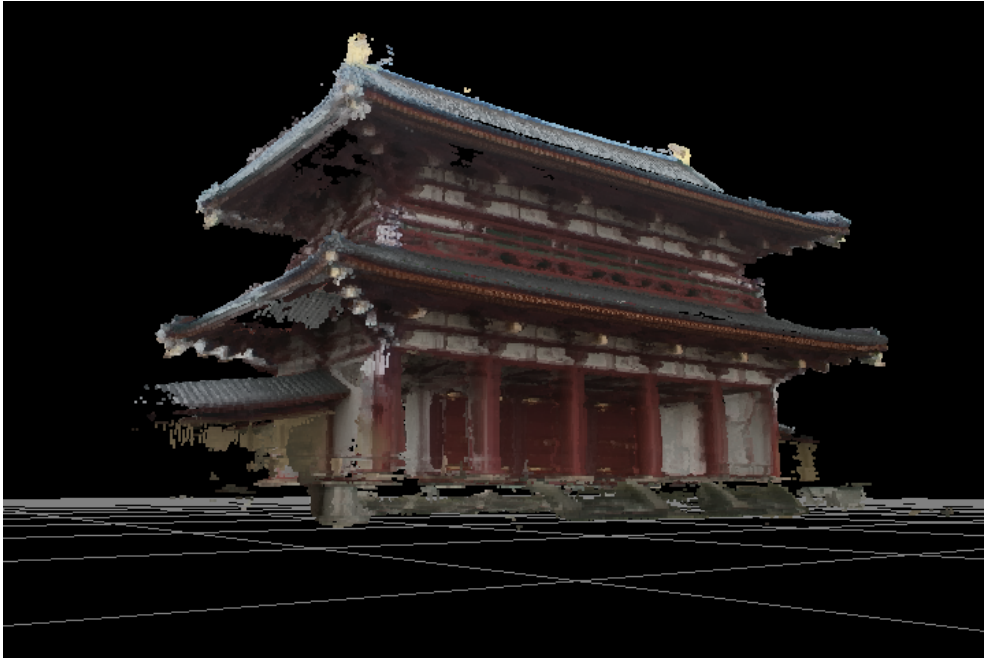
図 3.10 復元された三次元モデル (建物)





(b) 建物正面

図 3.10 復元された三次元モデル(建物)(つづき)



(c) 建物背面

図 3.10 復元された三次元モデル(建物)(つづき)



### 3.6.2 町並みを対象とした三次元モデル化 (実験 2)

本実験では一台のビデオカメラを三脚で車に固定し、車の進行方向に対して、(a) 前方、(b) 後方、にカメラを向けて同一の町並みを二つの動画像系列として撮影し、入力として用いた。得られた画像は、前方 500 枚と後方 405 枚である。図 3.11 に、得られた画像の一部を示す。これらの動画像に対して、2 章で述べた手法を用いてカメラパラメータの推定を行い、それぞれの動画像に対して図 3.12(a), (b) に示すカメラパラメータを得た。同図 (c) は、二つのカメラパラメータを同時に示したものである。また、図中の錘台は 30 フレーム毎のカメラの姿勢を表している。

以下では、入力とした動画像系列に対して、まず日照条件の変化の影響を削減し、次に奥行き画像の推定における SSD 算出時のウィンドウサイズについての検討を行なう。最後に、複数の動画像を用いて奥行き画像を推定し、ボクセル空間への統合を行なうことで三次元モデルを復元する。

#### (1) 日照条件の変化の影響の削減

二つの動画像間で、複数の対応点を画像上で指定することにより、3.4.1 節で述べた動画像間での色の変換係数を算出した。続いて、算出された色変換係数を用いることで色の補正を行い、日照条件の変化の影響の削減した。図 3.13 に、色が補正された前後の入力画像の例を示す。同図から、補正前には、図中 (b) の画像が図中 (a) の画像に対して暗いことが分かるが、補正後には、ほぼ等しい明るさに修正されていることが確認できる。

#### (2) SSD 算出のウィンドウサイズに関する予備実験

色の補正を行った二つの動画像系列を用いて、(a)  $1 \times 1$ 、(b)  $3 \times 3$ 、(c)  $5 \times 5$ 、および (d)  $3 \times 3$  (最下層のみ  $1 \times 1$ ) の異なる 4 つのウィンドウサイズで奥行き値の推定を行い、推定結果を比較した。図 3.14 に奥行き推定に用いた画像を、図 3.15(a), (b), (c), (d) に、それぞれのウィンドウサイズを用いて推定した奥行き画像を示す。奥行き画像の算出に要した時間は、それぞれ、260 秒 (a)、590 秒 (b)、1100 秒 (c)、230 秒 (d) である。

同図に示した (a), (b), (c), (d) および図 3.14 から、前項での実験と同様の傾向が見られるが、特に、ウィンドウサイズの大きな (b), (c) では、街灯の輪郭部分において、奥行き値が本来の形状に対して膨張するように推定され、誤りが発生して

いることが確認できる。これに対して、(a)、(d) では、共に街灯の輪郭部分の膨張は見られず、良好な結果を得ている。ただし、(a) では全体に誤った推定結果が多く見られる。

これらの結果から、町並みを対象とした実験においても、復元精度、計算時間ともに、(d) が優れており、前項での結果と同様に層によって異なるウインドウを用いた SSD の算出が、推定精度の向上および計算コストの削減に対して有効であると考えられる。

### (3) 三次元モデルの復元実験

(2) の結果から、色の補正を行った二つの動画像系列を用いて、画像ピラミッドの最下層で  $1 \times 1$ 、その他の層で  $3 \times 3$  のウインドウを用いて奥行き画像を推定した。図 3.16 に推定された奥行き画像の一部を示す。同図を含む奥行き画像から、奥行きの不連続な街路樹と背後の建物の間においても正しく奥行き値の推定が行われていることを確認した。しかし、ビルの窓ガラスのような鏡面性の部分や、鏡面成分の強い一部のビルの壁面において、奥行き値の推定に誤りが見られた。本実験においても、24 台の PC から成る PC クラスタ (Pentium4 1.8GHz, 1GB Memory) を用いて並列で奥行き画像を推定し、約 900 枚の奥行き画像を算出するのに要した時間は 87 分であった。

図 3.16 を含む約 900 枚の奥行き画像を、現実空間において一辺が 10cm のボクセルを用いて統合することで図 3.17 に示すモデルを得た。実験により、全体的には正しく町並みのモデルが構成されており、街路樹と建物の前後関係も正しく復元されたことから、複雑な屋外環境においても三次元モデル化が可能であることが確認できた。しかし、建物上部やオクルージョンにより隠蔽されるモデルの一部において、統合する画像の枚数が十分でないために建物の壁などに穴がみられた。また、テクスチャの無い一部のビルの壁面で穴が見られた。



第1フレーム



第330フレーム



第170フレーム



第500フレーム

(a) 前方



第1フレーム



第270フレーム



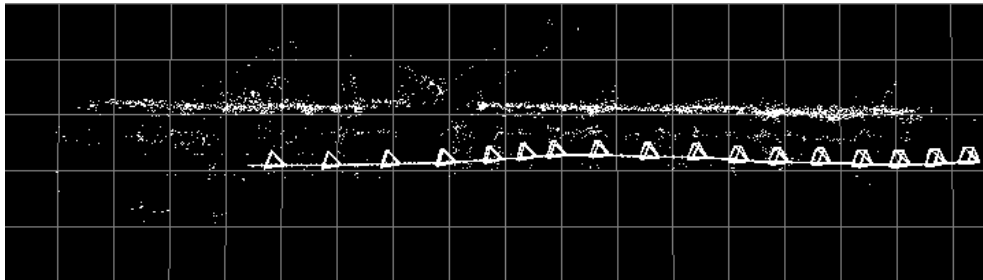
第135フレーム



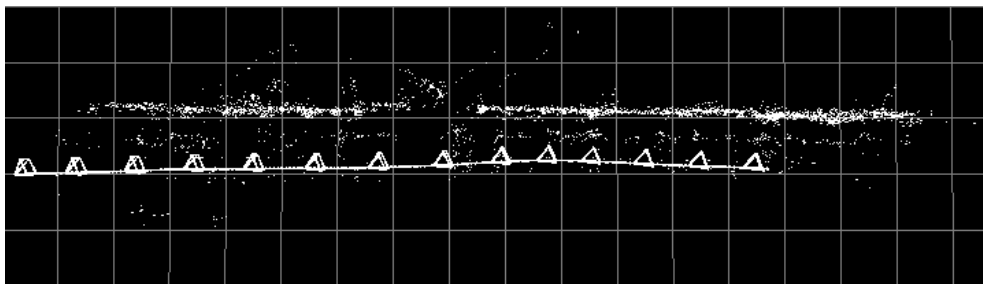
第405フレーム

(b) 後方

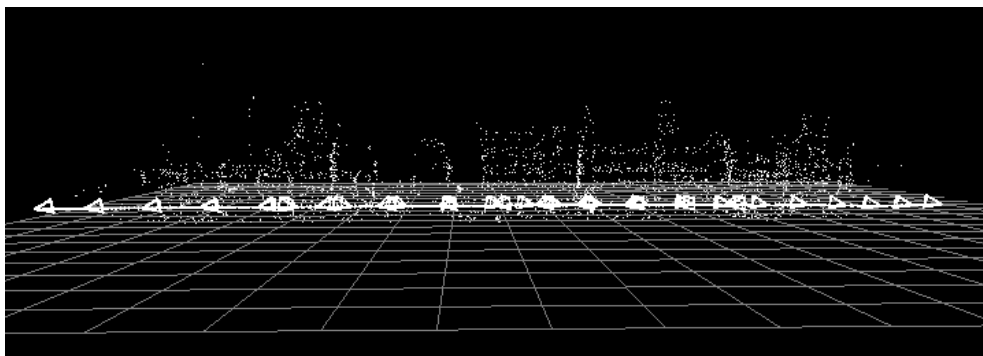
図 3.11 入力画像 (町並み)



(a) 前方(町並み上方)



(b) 後方(町並み上方)



(c) 前方+後方(町並み側方)

図 3.12 復元されたカメラパラメータと自然特徴点の三次元位置(町並み)



(a) 後方の画像



(b) 前方の画像 (補正前)

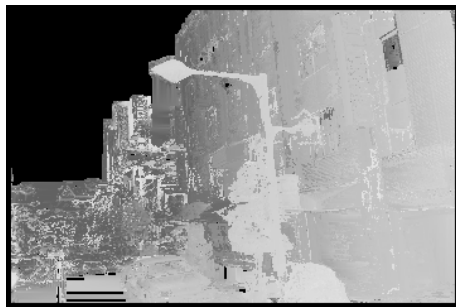


(c) 前方の画像 (補正後)

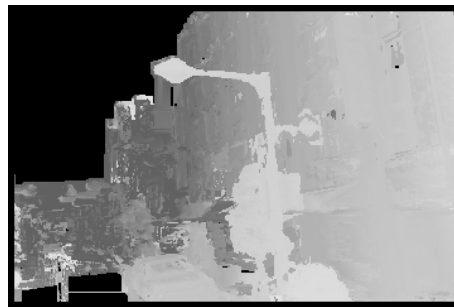
図 3.13 色の補正結果 (町並み)



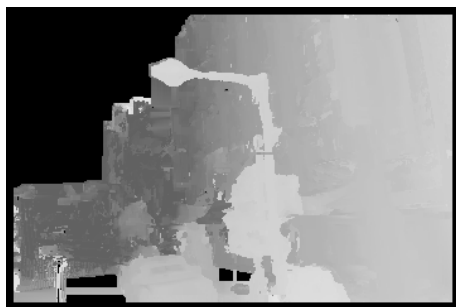
図 3.14 SSD 算出のウィンドウサイズによる奥行き推定結果の比較に用いた入力フレームの画像 (町並み)



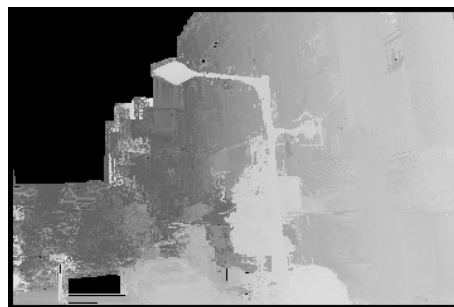
(a)  $1 \times 1$



(b)  $3 \times 3$



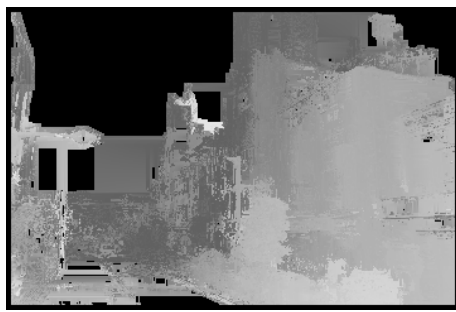
(c)  $5 \times 5$



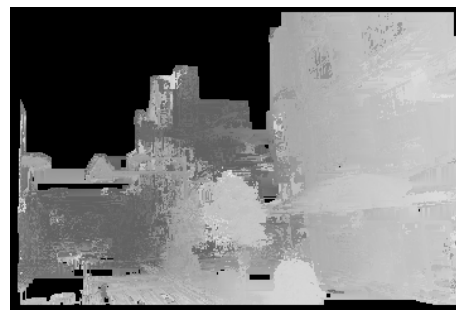
(d)  $3 \times 3$  (最下層のみ  $1 \times 1$ )

図 3.15 SSD 算出のウィンドウサイズによる奥行き画像の比較 (町並み)

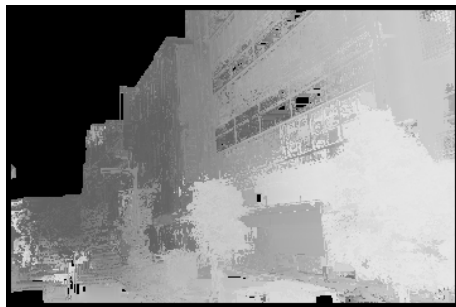




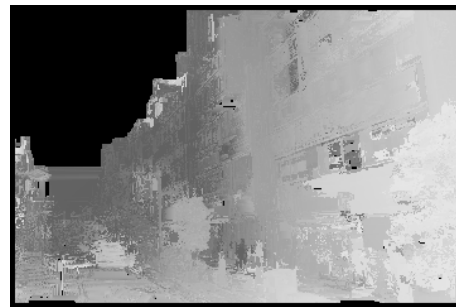
第 100 フレーム



第 300 フレーム

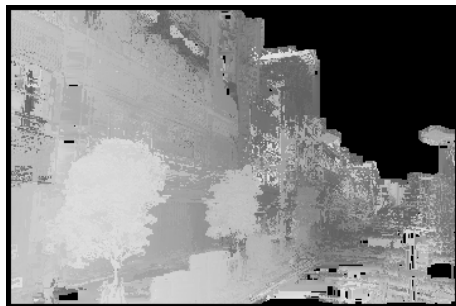


第 200 フレーム

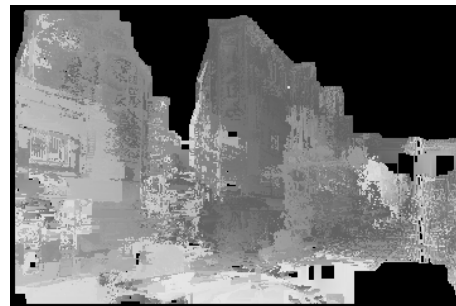


第 400 フレーム

(a) 前方



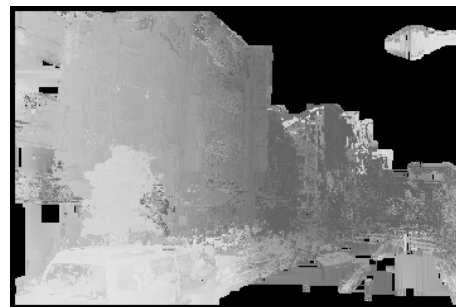
第 80 フレーム



第 240 フレーム



第 160 フレーム



第 320 フレーム

(b) 後方

図 3.16 奥行き画像の推定結果 (町並み)

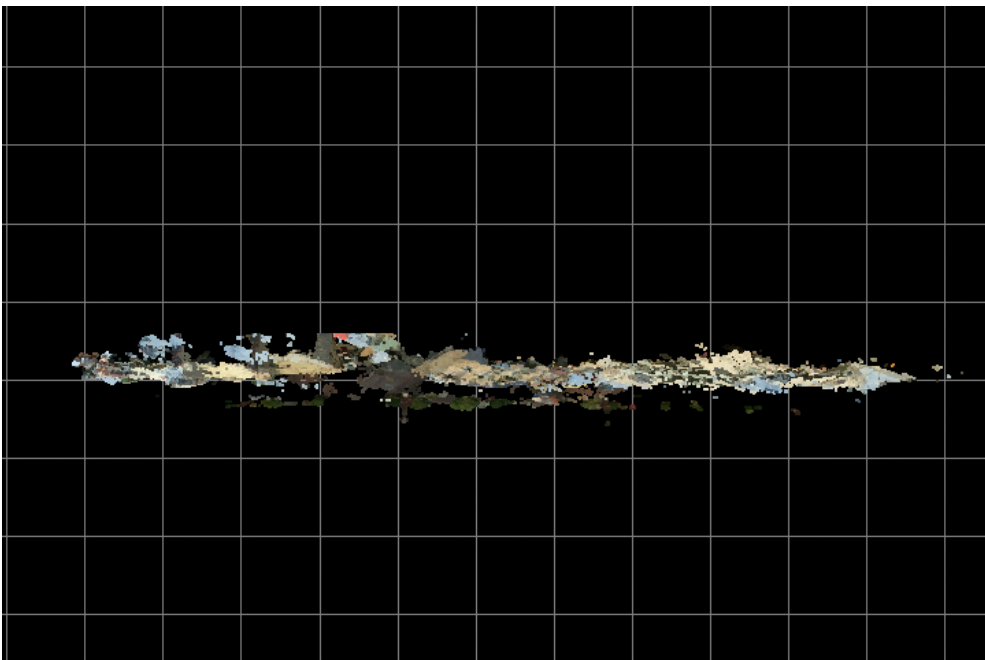
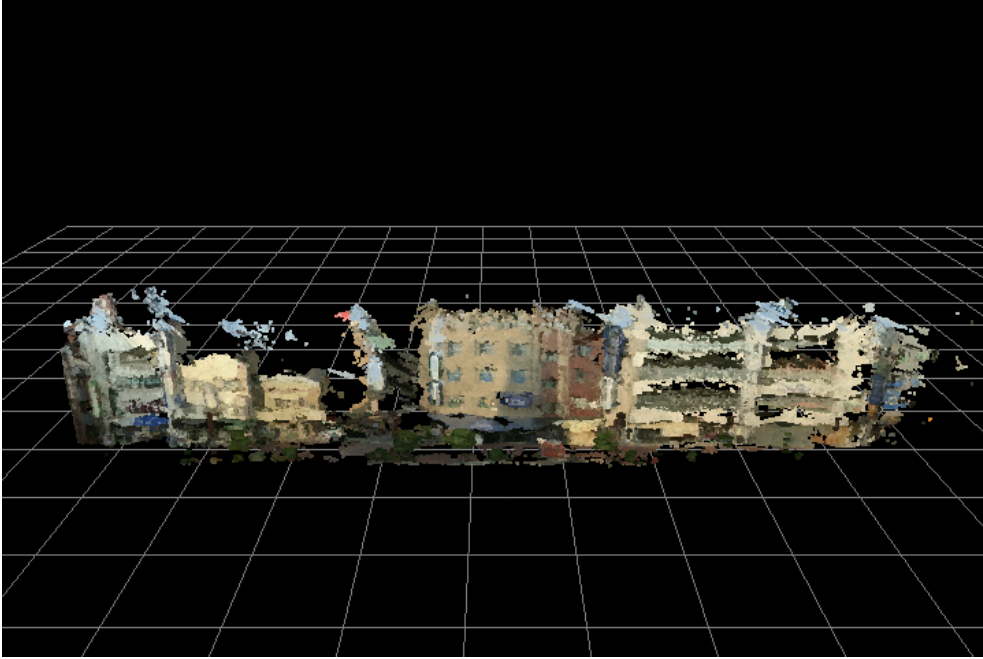


図 3.17 復元された三次元モデル(町並み)



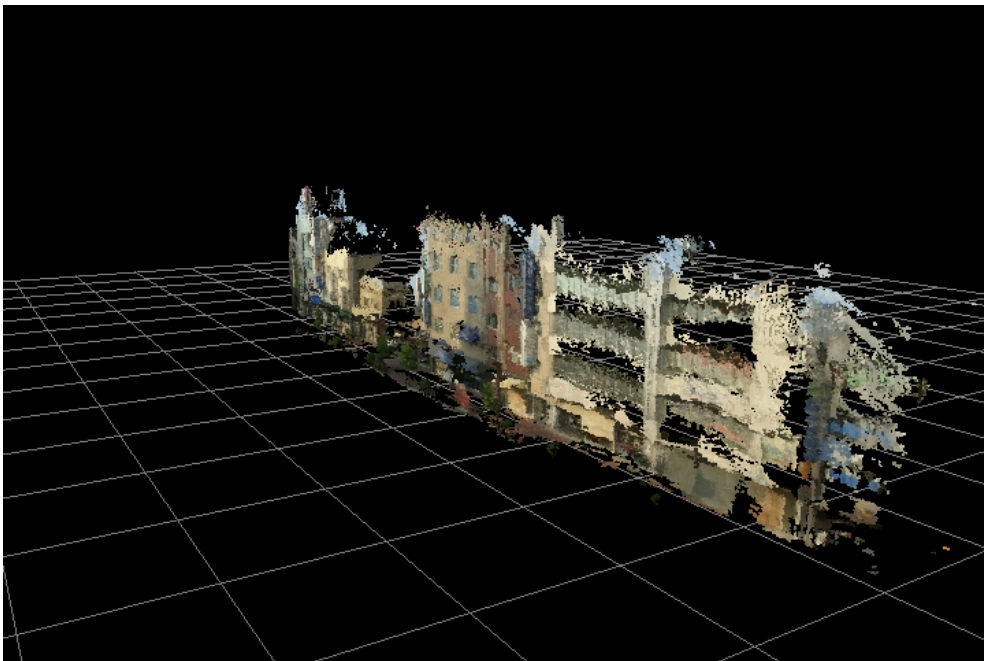
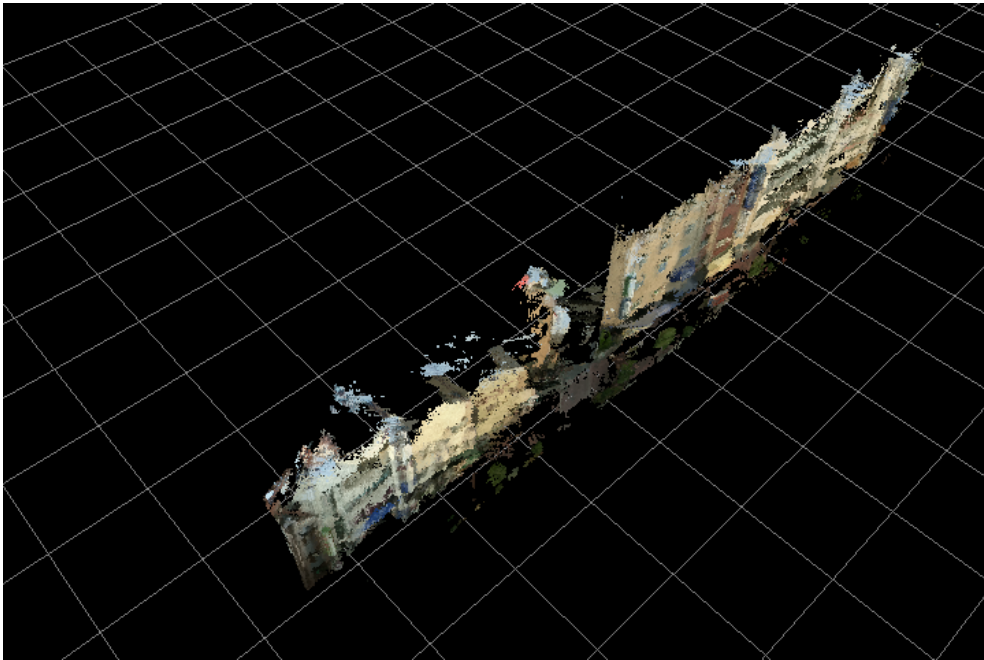


図 3.17 復元された三次元モデル(町並み) (つづき)

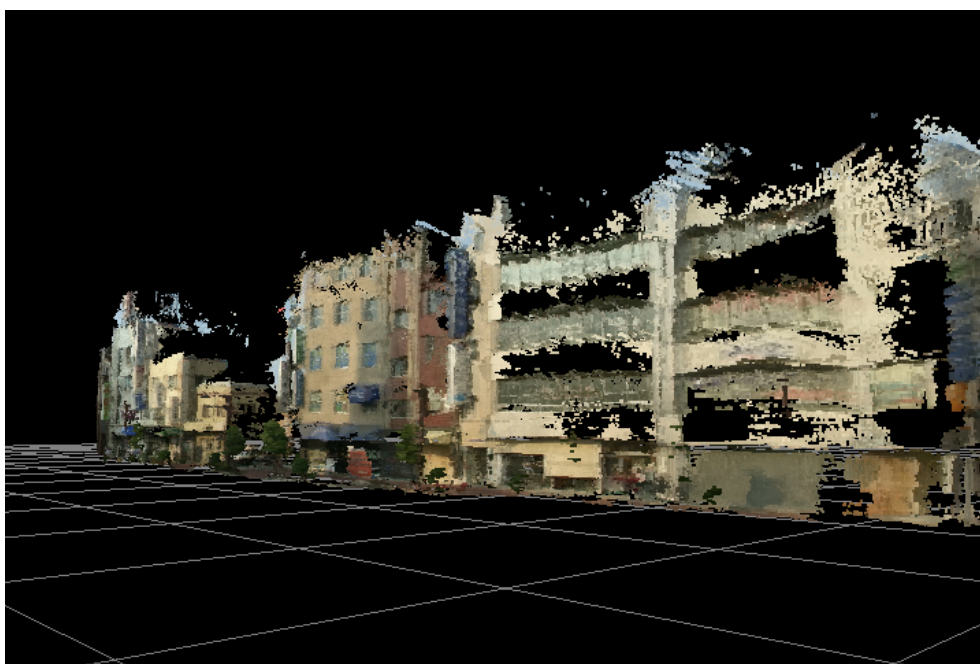
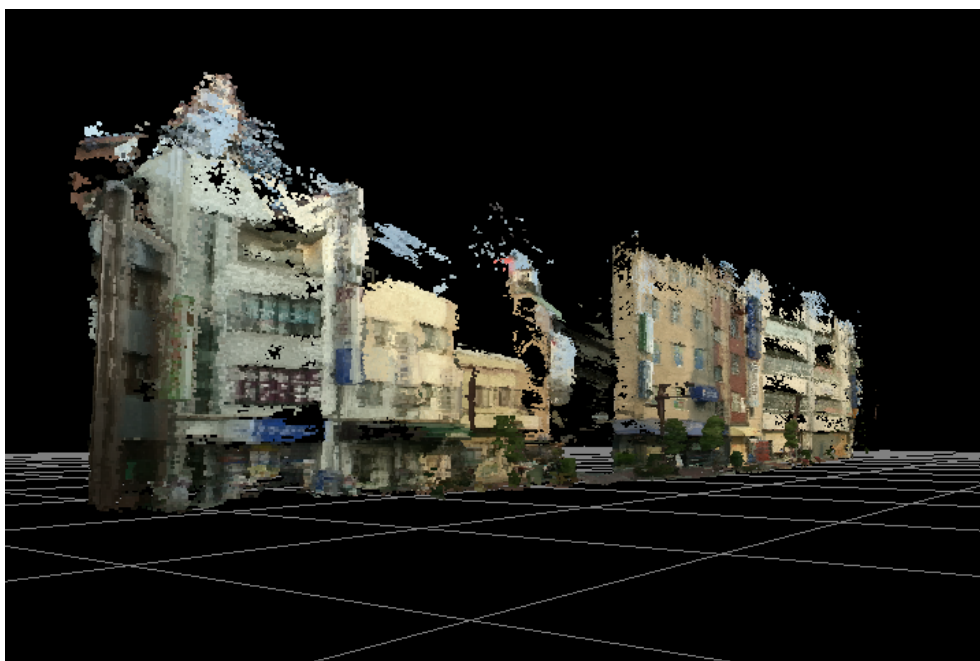


図 3.17 復元された三次元モデル(町並み) (つづき)

### 3.6.3 考察

前述した実験結果において、建物を対象とした実験では、建物の柱などの隠蔽関係が複雑となる部分に対してもおおむね正しく復元が行われていることが確認できた。また、町並みを対象とした復元においては、街路樹や街灯が多数配置された複雑な環境であったにもかかわらず、それらの前後関係などが正しく復元された。しかし、両実験共に、入力画像に写る枚数が少ない部分に関しては、復元されたモデルに穴が開くといった問題がみられた。また、町並みを対象とした実験においては、鏡面反射成分の強い一部の壁面や、窓ガラスのような鏡面性の物体の一部で奥行き画像が正しく復元されず、モデルの一部に穴が開くために、見た目の違和感が生じた。

入力画像に写る枚数が少ない部分に関しては、撮影に用いる動画像系列を増やすことで、復元精度の改善が可能であると考えられる。しかし、鏡面物体や撮影が困難な屋根などに関しては、動画像のみを入力として復元することが困難であるため、完全な三次元モデルを得るためには、これらを自動・手動で修正する手法の検討が必要であるといえる。

### 3.7 結言

本章では、カメラパラメータが推定された複数の動画像を入力とし、撮影対象の三次元モデルを復元する手法を提案した。本手法では、マルチベースラインステレオ法に対していくつかの拡張を行なうことで、多数のオクルージョンが発生する複雑な屋外環境においても、正しい奥行き情報を推定することが可能である。また、複数の動画像を統合的に用いて、奥行き情報を推定することにより、奥行き情報の推定精度を向上させている。

建物と町並みの復元実験においては、本手法が屋外環境のような複雑な環境に対しても、安定に奥行き情報を復元でき、またそれらを統合することにより三次元モデルを復元できることを確認した。しかし、特に町並みのような複雑で広域な環境においては、鏡面性の物体や入力画像に写る枚数が少ない部分に関して、奥行きが正しく推定されなかったため、部分的にモデルが復元されず三次元モデルに穴がみられた。

入力画像に写る枚数が少ない部分に関しては、入力として用いる動画像を増やすことや、全方位型のカメラを用いることで、推定精度を向上させることが可能であると考えられる。しかし、鏡面物体が復元されないという問題に関しては、本手法のように画像のみを用いた三次元復元手法によって解決することは難しいといえる。このため、完全な三次元モデルを得るためには、これらを自動・手動で修正する手法の検討が必要であるといえる。

## 4. 結論

### 4.1 まとめと考察

本論文では、広範囲の屋外環境を少ない制約・コストで三次元モデル化する手法について提案した。従来から、屋内のような多数の制約条件やセンサ・マーカを用いることのできる環境では、様々な三次元モデル化の手法が提案され実用化されてきたが、屋外環境のような複雑で広域な環境への適用は困難であった。

これに対して、本研究では、撮影の手間や撮影対象に対する制約が少ない動画像による手法を用いて屋外環境の三次元モデル化を行う手法を提案した。本手法は、まず、屋外環境を撮影した複数の動画像系列のカメラパラメータを推定し、次に、各入力画像に対して奥行き画像を推定し、ボクセル空間への統合を行うことで、屋外環境の三次元モデル化を行う。

動画像のカメラの移動パラメータの推定においては、従来手法の問題であった復元範囲や誤差の累積に関する問題に対して、三次元測量機材であるトータルステーションを用いて環境の一部を計測し、基準マーカとして全体的な最適化処理に用いることで解決を図った。また、実際の屋外環境を単眼のカメラを用いて複数の動画像系列に分けて撮影し、本手法を適用することによって、カメラパラメータの復元が可能であることを確認した。定量的な評価実験では、100m程度の移動を伴う屋外環境の撮影においては、カメラ位置に関して平均誤差 100mm 程度、カメラの光軸方向に関して平均誤差 0.1°程度でカメラパラメータを得ることが可能であることを確認した。加えて、Match Moveを行うことによって景観シミュレーションなどの応用アプリケーションにおいても利用が可能であることを示した。

奥行き画像の推定による三次元モデルの復元においては、カメラパラメータが推定された複数の動画像系列に対して、拡張マルチベースラインステレオ法を用いる手法を提案した。本手法は、従来のマルチベースライン法を拡張し、オクルージョンの考慮や、日照条件の削減を行うことで、より安定な奥行き画像の推定を実現する。実験では、カメラパラメータが推定された複数の動画像を用いて建物と町並みの三次元モデル化を行った。建物を対象とした実験では、建物の柱などの隠蔽関係が複雑となる部分に対してもおおむね正しく復元が行われていることが

確認できた。しかし、入力画像に写る枚数が少ない部分に関しては、復元されたモデルに穴が開くといった問題がみられた。また、町並みを対象とした復元においては、街路樹や街灯が多数配置された複雑な環境であったにもかかわらず、それらの前後関係などが正しく復元された。しかし、鏡面反射成分の強い一部の壁面や、窓ガラスのような鏡面性の物体の一部で奥行き画像が正しく復元されず、モデルの一部に穴が開くために、見た目の違和感が生じた。

これらの結果から、現時点では仮想空間内でのウオークスルーのようにモデル自体を見せることを目的としたアプリケーションで、直接利用することは困難であると思われる。ただし、屋外環境の概形はおおむね正しく復元されているため、ツールを用いて手動でモデルの修正を行うことは比較的容易であると考えられる(付録B参照)。また、鏡面物体や撮影が困難な屋根などは、動画像のみを入力として復元することが困難であるため、完全な三次元モデルを得るためには、これらを自動・手動で修正する手法の検討が必要である。

## 4.2 今後の展望

本論文では、複雑な屋外環境を自動で復元する手法を提案したが、本手法をより実用的な三次元モデルの復元手法に発展させるため、以下の手法を検討することが考えられる。

### 全方位型カメラの利用

本研究で用いたカメラは、一般的な単眼のカメラであるために、撮影できる範囲が限られており、オクルージョンによるモデルの欠落なしに、町並みの両側面を復元することを考えた場合には、カメラの向き・経路を変えて複数回の撮影を行う必要がある。

これに対し、入力センサとして全方位型のカメラシステムを用いれば、一回の撮影によって町並みの両側面を復元可能であり、また全方位を観測しながら移動できるため、オクルージョンにも強いという利点がある。更に、全方向の自然特徴点を同時に利用できるため、カメラパラメータの推定精度の向上も期待できる。

また従来は、全方位カメラに関して、解像度の不足が問題となっていたが、近年、図 4.1 に示すような、高解像度の全方位画像を取得可能な、マルチカメラシステム [池田 02] や Hi-Vision 型の全方位カメラ [山澤 01] などが開発されているため、これらを用いた実装による復元精度の向上と、画像取得コストの低下が期待できる。

#### 能動的な三次元復元手法の併用

近年、光レーザに基づくレンジファインダを用いて、屋外環境を計測する手法が盛んに研究されている。光レーザに基づくレンジファインダでは、奥行き情報に対する画素の色を決定することが問題とされていたが、1 pixel CCD を計測装置に内蔵することで、奥行き情報と同期したテクスチャ情報を獲得することが可能となってきた。

実際に Riegl 社製レーザレンジファインダ [Rie02] を用いて屋外環境を計測し、仮想空間内で点群プロットすることによって可視化した例を図 4.2 に示す。このような装置は、100m 程度の範囲内での計測精度に関してはほぼ問題がないといえるが、移動を伴った撮影ができないため、同図中に見られるようにオクルージョンによるモデルの欠損がおこり、各計測地点間での位置あわせが必要となるという問題が残されている。

一方、本論文で提案した手法では、動きながらのデータの獲得ができるために、比較的オクルージョンに強く、また撮影位置の推定が可能である。そこで、能動的な計測データと動画像による復元データを補完的に用いることで、より完成度の高い三次元モデルの復元が期待できる。

また、本手法で復元した奥行き画像・三次元モデルとカメラパラメータを用いた応用として以下のものが考えられる。

#### 物体による隠蔽を考慮した蓄積型景観シミュレーション

提案した手法では、動画像の各フレームにおいて、奥行き画像を復元することができる。これらを Match Move に利用することで、CG と現実物体の隠蔽関係を正しく再現することが可能となる。ただし、奥行き画像の推定には



(a) 全方位マルチカメラシステム (b)Hi-Vision 型の全方位カメラ

図 4.1 高解像度の全方位画像を取得可能な全方位センサ

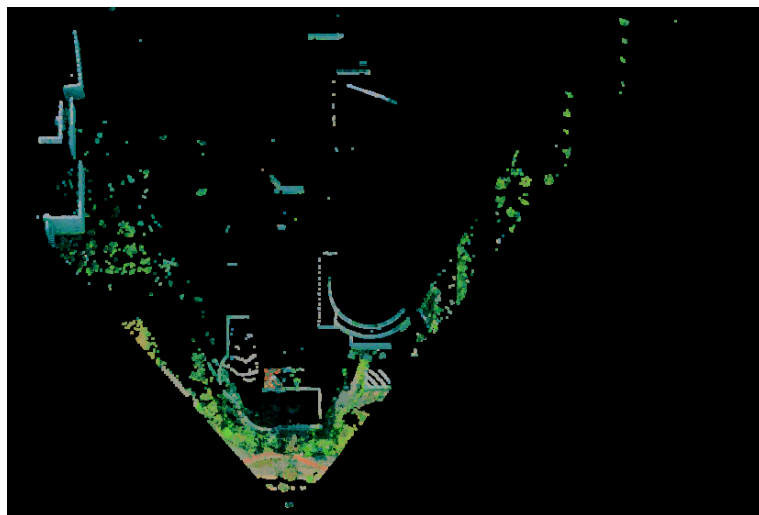
動画像全体を用いるために、リアルタイムで景観リミュレーションに応用することはできず、あくまでも蓄積型のシミュレーションとなる。

#### 三次元モデルをマーカとするカメラの移動パラメータの推定

本手法で復元された三次元モデルは、幾何学的な形状に加えてテクスチャ情報を保持している。この三次元モデル自体を三次元形状が既知のマーカとみなせば、対応する現実環境を撮影することによって、逆にカメラ位置を推定することが可能となる。このような技術は、ウェアラブル拡張現実感システム等のための、リアルタイムでのカメラ位置推定手法として利用できると考えられる。

このように、本研究によって得られる成果は、様々な分野に応用が可能である。今後は、上述した三次元モデルの復元精度向上と、それらを用いた応用分野についての検討が考えられる。





(a) 上方



(b) 側方

図 4.2 光レーザに基づくレンジファインダによる計測データの可視化例

## 謝辞

本研究の全過程を通して、懇切なる御指導、御鞭撻を賜った視覚情報メディア講座 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、有益な御助言と御鞭撻を賜っ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(現、大阪大学サイバーメディアセンター教授)に深く感謝致します。

そして、本研究を通じて、有益な御助言を頂いた視覚情報メディア講座 山澤 一誠助教授に厚く感謝致します。

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視覚情報メディア講座 神原 誠之助手に心から感謝致します。

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、本研究の実験・データ計測において、多大な協力を頂いた池田 聖氏、視覚情報メディア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視覚情報メディア講座 事務補佐員 北川 知代女史、元ソフトウェア基礎講座 事務補佐員 福永 博美女史、に厚く感謝致します。

## 参考文献

- [AL87] N. Ayache and F. Lustman, “Fast and Reliable Passive Trinocular Stereo Vision,” *Proc. IEEE Int. Conf. on Computer Vision*, pp. 422-426, 1987.
- [Ard89] G. Ardeshir, “選択的探索によるステレオ対応づけ”, 情報処理学会 研究報告, CV-61-4, 1989.
- [BF97] R. Benjemma and F. Schmitt, “Fast Global Registration of 3D Sampled Surface Using a Multi-z-buffer Technique,” *Proc. IEEE Int. Conf. on Recent Advance in 3-D Digital Imaging and Modeling*, pp. 113-120, 1997.
- [BZM97] P. Beardsley, A. Zisserman and D. Murray, “Sequential Updating of Projective and Affine Structure from Motion,” *Int. Journal of Computer Vision*, Vol. 23, No. 3, pp. 235-259, 1997.
- [Cyb93] Cyberware, “Cyberware 3030 RGB,”  
<http://www.cyberware.com/products/psInfo.html>, 1993.
- [Cyr00] Cyra, “Cyrax 2500,”  
[http://www.cyra.com/products/cyrax2500\\_description.html](http://www.cyra.com/products/cyrax2500_description.html), 2000.
- [Dis02] Discreet, “3ds max,”  
<http://www.discreet.com/products/3dsmax/>, 2002.
- [HM98] S. Hiura and T. Matsuyama, “Depth Measurement by The Multi-focus Camera,” *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 953-959, 1998.
- [Hor75] B. K. P. Horn, “Obtaining Shape from Shading Information,” in *The Psychology of Computer Vision*, pp. 115-155. McGraw-Hill, 1975.
- [Hor86] B. K. P. Horn, *Robot Vision*. The MIT Press, 1986.

- [HS88] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Alvey Vision Conf.*, pp. 147-151, 1988.
- [Ike81] K. Ikeuchi, "Determining Surface Orientation of Specular Surfaces by using The Photometric Stereo Method," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 3, No. 6, pp. 661-669, 1981.
- [LJ77] R. A. Lewis and A. R. Johnston, "A Scanning Laser Rangefinder for a Robotic Vehicle," *Proc. Int. Joint Conf. on Artificial Intelligence*, pp. 762-768, 1977.
- [MIN02] MINOLTA, "VIVID 910," <http://www.minolta-rio.com/vivid/>, 2002.
- [MK85] V. J. Milenkovic and T. Kanade, "Trinocular Vision Using Photometric and Edge Orientation constraints," *Proc. DARPA Image Understanding Workshop*, pp. 163-175, 1985.
- [MK98] 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.
- [MO99] 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. IEEE Int. Conf. on Computer Vision*, pp. 411-417, 1999.
- [MSO96] A. Matsumoto, H. Saito and S. Ozawa, "3-D Reconstruction of Skin Surface from Photometric Stereo Images with Specular and Inter Reflections," *Proc. IEEE Int. Conf. on Control, Automation, Robotics, and Vision*, Vol. 2, pp. 778-782, 1996.
- [MSY96] T. Masuda, K. Sakaue and N. Yokoya, "Registration and Integration of Multiple Range Images for 3-D Models Construction," *Proc. IEEE*

- Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 879-883, 1996.
- [Nay92] S. K. Nayar, "Shape from Focus System," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 302-308, 1992.
- [NBD77] D. Nitzan, A. E. Brain and R. O. Duda, "The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis," *Proc. of the IEEE*, Vol. 65, No. 2, pp. 206-219, 1977.
- [Neu97] P. Neugebauer, "Geometric Cloning of 3D Objects via Simultaneous Registration of Multiple Range Images," *Proc. IEEE Int. Conf. on Shape Modeling and Application*, pp. 130-139, 1997.
- [New02] Newtek, "LIGHTWAVE 3D," <http://www.lightwave3d.com/>, 2002.
- [NN94] S. K. Nayar and Y. Nakagawa, "Shape from Focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 8, pp. 824-831, 1994.
- [NS92] H. N. Nair and C. V. Stewart, "Robust Focus Ranging," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 309-314, 1992.
- [NW97] W. Niem and J. Wingbermhle, "Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera," *Proc. IEEE Int. Conf. on Recent Advances in 3-D Digital Imaging and Modeling*, pp. 173-180, 1997.
- [NWN96] S. K. Nayar, M. Watanabe and M. Noguchi, "Real-time Focus Range Sensor," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 12, pp. 1186-1198, 1996.

- [OK85] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter- Scanline Search Using Dynamic Programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, No. 2, pp. 139-154, 1985.
- [OK93] M. Okutomi and T. Kanade, "A Multiple-baseline Stereo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 353-363, 1993.
- [PK93] J. Poleman and T. Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery," *Tech. Rep. CMU-CS-93-219, Carnegie-Mellon Univ.*, 1993.
- [PK+00] 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.
- [Poi01] PointGreyResearch, "Digiclops,"  
<http://www.ptgrey.com/products/digiclops/index.html>, 2001.
- [Pul99] K. Pulli, "Multiview Registration for Large Data Sets," *Proc. IEEE Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 160-168, 1999.
- [QK96] L. Quan and T. Kanade, "A Factorization Method for Affine Structure from Line Correspondence," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 803-808, 1996.
- [RC97] A. N. Rajagopalan and S. Chaudhuri, "Optimal Selection of Camera Parameters for Recovery of Depth from Defocused Images," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 219-224, 1997.
- [Rie02] Riegl, "LMS-Z360,"  
[http://www.riegl.co.at/lmsz360/e\\_lmsz360.htm](http://www.riegl.co.at/lmsz360/e_lmsz360.htm), 2002.

- [RW00] G. Roth and A. Whitehead, "Using Projective Vision to Find Camera Positions in an Image Sequence," *Proc. Int. Conf. on Vision Interface*, pp. 87-94, 2000.
- [SK94] 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.
- [SK+01] T. Sato, M. Kanbara, H. Takemura and N. Yokoya, "3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features," *Proc. Int. Conf. on Vision Interface*, pp. 157-164, 2001.
- [SM+02] T. Sato, M. Kanbara, N. Yokoya and H. Takemura, "Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera," *Int. Journal of Computer Vision*, Vol. 47, No. 1-3, pp. 119-129, 2002.
- [SMB98] C. Schmid, R. Mohr and C. Bauckhage, "Comparing and Evaluating Interest Points," *Proc. IEEE Int. Conf. on Computer Vision*, pp. 230-235, 1998.
- [SY+99] 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.
- [TK91] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," *Tech. Rep. CMU-CS-91-132, Carnegie-Mellon Univ.*, 1991.
- [TK92] 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.

- [Tsa86] R. Y. Tsai, "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 364-374, 1986.
- [Woo80] R. J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *Journal of Optical Engineering*, Vol. 19, No. 1, pp. 139-144, 1980.
- [YSK99] N. Yokoya, T. Shkunaga and M. Kanbara, "Passive Range Sensing Techniques: Depth from Images," *IEICE Trans. on Information and Systems*, Vol. E82-D, No. 3, pp. 523-533, 1999.
- [ZG98] L. X. Zhou and W. K. Gu, "3D Model Reconstruction by Fusing Multiple Visual Cues," *Proc. IAPR Int. Conf. on Pattern Recognition*, pp. 640-642, 1998.
- [Zhe94] 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.
- [ZS00] H. Zhao and R. Shibasaki, "Reconstruction of Textured Urban 3D Model by Fusing Ground Based Laser Range and CCD Images," *IEICE Trans. on Information and Systems*, Vol. E83-D, No. 7, pp. 1429-1440, 2000.
- [池田 02] 池田 聖, 佐藤 智和, 横矢 直和, "全方位型マルチカメラシステムのキャリブレーション", 情報科学技術フォーラム (FIT) 情報技術レターズ, Vol. 1, pp. 149-150, 2002.
- [井口 90] 井口 征士, 佐藤 宏介, 三次元画像計測. 昭晃堂, 1990.
- [植芝 97] 植芝 俊夫, 富田 文明, "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", 情報処理学会 研究報告, CVIM107-1, 1997.



- [上原 00] 上原 将文, 塩崎 剛志, 全 炳東, “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.
- [浮田 98] 浮田 宗伯, 尺長 健, “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”, 電子情報通信学会 技術研究報告, PRMU97-276, 1998.
- [岡谷 00] 岡谷 貴之, 出口 光一郎, “視点の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.
- [栗田 00] 栗田 多喜夫, 坂上 勝彦, “ロバスト統計とその画像理解への応用”, 画像の認識・理解シンポジウム (MIRU 2000) 講演論文集, Vol. I, pp. 65-70, 2000.
- [出口 90] 出口 光一郎, “射影幾何学による PnP カメラ補正問題の統一的解法”, 情処シンポジウム, Vol. 90, pp. 41-50, 1990.
- [出口 93] 出口 光一郎, “カメラキャリブレーション手法の最近の動向”, 情処研報, CV-82-1, 1993.
- [出口 97] 出口 光一郎, “透視投影条件の下での因子分解法による多視点画像からの形状復元”, 情報処理学会 研究報告, CVIM106-6, 1997.
- [戸所 98] 戸所 信樹, 中村 納, “単一曲面物体の輪郭線情報と反射率分布図の関係に基づく3次元画像計測”, 電子情報通信学会論文誌 (D-II), Vol. J81-D-II, No. 12, pp. 2718-2730, 1998.
- [佐藤 85] 佐藤 宏介, 井口 征士, “空間コード化による距離画像入力”, 電子情報通信学会論文誌 (D), Vol. J68-D, No. 3, pp. 369-375, 1985.
- [佐藤 02] 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”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 275-282, 2002.

- [島村 00] 島村 潤, 山澤 一誠, 竹村 治雄, 横矢 直和, “全周実画像とCGモデルの合成による仮想環境の構築”, 画像の認識・理解シンポジウム (MIRU 2000) 講演論文集, Vol. II, pp. 367-372, 2000.
- [辰野 96] 辰野 由美子, 鈴木 哲, 横矢 直和, 岩佐 英彦, 竹村 治雄, “全周計測距離データを用いた顔基本6表情の解析と合成”, テレビジョン学会誌, Vol. 50, No. 10, pp. 1474-1481, 1996.
- [塚本 00] 塚本 壮輔, 呉 本亮, 古賀 和利, 三池 秀敏, “階層化位相シフト法による高精度な奥行き計測”, 画像の認識・理解シンポジウム (MIRU 2000) 講演論文集, Vol. I, pp. 185-190, 2000.
- [西野 02] 西野 恒, 池内 克史, “大規模距離画像群の頑健な同時位置合せ”, 電子情報通信学会論文誌 (D-II), Vol. J85-D-II, No. 9, pp. 1413-1424, 2002.
- [藤井 00] 藤井 博文, 神原 誠之, 竹村 治雄, 横矢 直和, “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”, 電子情報通信学会 技術研究報告, MVE99-59, 2000.
- [増田 02] 増田 健, “符号付距離場の一致による複数距離画像からの形状モデル生成”, 画像の認識・理解シンポジウム (MIRU 2002) 講演論文集, Vol. I, pp. 1-10, 2002.
- [町田 01] 町田 貴史, “複数の照明条件の組み合わせによる物体の表面反射特性の密な推定”, 電子情報通信学会論文誌 (D-II), Vol. J84-D-II, No. 8, pp. 1873-1881, 2001.
- [松川 00] 松川 昌洋, 神原 誠之, 山澤 一誠, 竹村 治雄, 横矢 直和, “GPSと三軸角度センサを用いた多視点ステレオによる屋外環境構造の再構成”, 電子情報通信学会 技術研究報告, PRMU99-266, 2000.
- [山口 91] 山口 証, 中山 収文, 白井 良明, 浅田 稔, “信頼性の高い対応を優先した多段階ステレオ法”, 電子情報通信学会論文誌 (D-II), Vol. J74-D-II, No. 7, pp. 918-925, 1991.

- [山澤 01] 山澤 一誠, 竹村 治雄, 横矢 直和, “全方位HD カメラを用いたテレプレゼンシステム”, 第6回パターン計測シンポジウム講演論文集, pp. 15-20, 2001.
- [横光 99] 横光 澄男, 大隈 隆史, 竹村 治雄, 横矢 直和, “多視点ステレオ画像を用いた屋外環境構造の再構築”, 電子情報通信学会 技術研究報告, PRMU98-250, 1999.
- [横矢 93] 横矢 直和, “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”, 電子情報通信学会論文誌 (D-II), Vol. J76-D-II, No. 8, pp. 1667-1675, 1993.
- [盧 99] 盧 存偉, 井口 征士, “パターン光投影に基づく3次元計測における縞次数の自動判定”, 電子情報通信学会 技術研究報告, PRMU99-193, 1999.
- [渡辺 02] 渡辺 真太郎, 宮島 耕治, 井上 潮, 中村 太一, “ステレオマッチングを応用した撮影の位置と時期が異なる航空写真からの建造物の変化抽出”, 電子情報通信学会論文誌 (D-II), Vol. J85-D-II, No. 10, pp. 1531-1540, 2002.

## 研究業績

### 学術論文誌

1. T. Sato, M. Kanbara, N. Yokoya and H. Takemura, “Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera”, *Int. Journal of Computer Vision*, Vol. 47, No. 1-3, pp. 119–129, April 2002. (第 2, 3 章に関連)
2. 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”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 275–282, June 2002. (第 3 章に関連)
3. 山本 光重, 佐藤 智和, 横矢 直和, “実写映像を利用した花火演出支援システム FirePainter”, 日本バーチャルリアリティ学会論文誌, Vol. 7, No. 3, pp. 385–392, October 2002.

### レター

1. 佐藤 智和, 神原 誠之, 横矢 直和, “複数の動画像系列を用いたマルチベースラインステレオによる屋外環境の三次元復元”, 情報科学技術フォーラム (FIT) 情報技術レターズ, Vol. 1, pp. 151–152, September 2002. (第 3 章に関連)
2. 池田 聖, 佐藤 智和, 横矢 直和, “全方位型マルチカメラシステムのキャリブレーション”, 情報科学技術フォーラム (FIT) 情報技術レターズ, Vol. 1, pp. 127–128, September 2002.

### 国際会議

1. T. Sato, M. Kanbara, H. Takemura and N. Yokoya, “3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features”, *Proc. 14th Int. Conf. on Vision Interface 2001 (VI2001)*, pp. 157–154, June 2001. (第 2 章に関連)
2. T. Sato, M. Kanbara, N. Yokoya and H. Takemura, “Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held

- Video Camera”, *Proc. IEEE Workshop on Stereo and Multi-Baseline Vision*, pp. 57–61, December 2001. (第 3 章に関連)
3. T. Sato, M. Kanbara, N. Yokoya and H. Takemura, “Dense 3-D Reconstruction from a Monocular Image Sequence by Estimating Camera Motion Parameters”, *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR2001)*, Technical Sketches, December 2001. (第 3 章に関連)
  4. T. Sato, M. Kanbara, N. Yokoya and H. Takemura, “3-D Modeling of an Outdoor Scene by Multi-baseline Stereo Using a Long Sequence of Images”, *Proc. 16th IAPR Int. Conf. on Pattern Recognition (ICPR2002)*, Vol. III, pp. 581-584, August 2002. (第 3 章に関連)
  5. S. Ikeda, T. Sato and N. Yokoya, “Calibration Method for an Omnidirectional Multicamera System”, *Proc. SPIE Electronic Imaging*, Vol. 5006-61, January 2003.

## 研究会・シンポジウム

1. 佐藤 智和, 神原 誠之, 竹村 治雄, 横矢 直和, “単眼動画像からのマーカと自然特徴点の追跡による三次元復元”, 画像電子学会第 182 回研究会講演予稿, pp. 29–34, November 2000. (第 2 章に関連)
2. 佐藤 智和, 神原 誠之, 竹村 治雄, 横矢 直和, “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”, 電子情報通信学会 技術研究報告, PRMU2000-144, December 2000. (第 2 章に関連)
3. 佐藤 智和, 神原 誠之, 竹村 治雄, 横矢 直和, “マーカと自然特徴点の追跡を用いた単眼動画像からの三次元復元”, 日本バーチャルリアリティ学会 研究報告 (サイバースペースと仮想都市研究会), Vol. 5, No. 2, pp. 5–10, January 2001. (第 2 章に関連)
4. 佐藤 智和, 神原 誠之, 横矢 直和, “視点位置推定による動画像からの屋外環境の三次元モデル化”, 画像電子学会第 190 回研究会講演予稿, pp. 49–55, November 2001. (第 3 章に関連)

5. 山本 光重, 佐藤 智和, 横矢 直和, “実写映像を利用した花火演出支援システム”, 電子情報通信学会 技術研究報告, IE2001-199, January 2002.
6. 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”, 電子情報通信学会 技術研究報告, PRMU2001-223, February 2002. (第 3 章に関連)
7. 山本 光重, 佐藤 智和, 横矢 直和, “実写映像を利用した花火演出支援システム FirePainter”, 画像の認識・理解シンポジウム (MIRU2002) 講演論文集, Vol. II, pp. 231-238, August 2002.
8. 佐藤 智和, 神原 誠之, 横矢 直和, 竹村 治雄, “マーカと自然特徴点の追跡による動画像からのカメラパラメータの推定と三次元復元”, 画像の認識・理解シンポジウム (MIRU2002) 講演論文集, Vol. II, pp. 7-12, August 2002. (第 2 章に関連)
9. 池田 聖, 佐藤 智和, 横矢 直和, “全方位型マルチカメラシステムによるパノラマ動画像の生成”, 電子情報通信学会 技術研究報告, PRMU2002-154, December 2002.

## 全国大会

1. 佐藤 智和, 神原 誠之, 竹村 治雄, 横矢 直和, “透視投影モデルにおける単眼動画像からの三次元復元”, 2000 年電子情報通信学会情報ソサイエティ大会講演論文集, No. D-12-49, September 2000. (第 2 章に関連)
2. 山本 光重, 佐藤 智和, 横矢 直和, “IBR による描画を利用した花火演出支援システム”, 日本バーチャルリアリティ学会第 6 回大会論文集, pp. 15-16, September 2001.
3. 佐藤 智和, 神原 誠之, 横矢 直和, 竹村 治雄, “視点位置推定に基づく動画像からの屋外環境の密な三次元復元”, 2001 年電子情報通信学会情報ソサイエティ大会講演論文集, No. D-12-43, September 2001. (第 3 章に関連)
4. 佐藤 智和, 神原 誠之, 横矢 直和, 竹村 治雄, “視点位置推定に基づく動画像からの密な屋外シーンの再構築”, 平成 13 年電気関係学会関西支部連合大会講演論文集, No. G17-16, November 2001. (第 3 章に関連)

5. 池田 聖, 佐藤 智和, 横矢 直和, “全方位型マルチカメラシステムのキャリブレーションによる全天球動画像の生成”, 平成 14 年電気関係学会関西支部連合大会講演論文集, No. G13-27, November 2002.

## 解説

1. 佐藤 智和, 神原 誠之, 横矢 直和, 竹村 治雄, “動画像からの三次元復元 マルチベースラインステレオ法を用いた屋外環境の三次元モデル化”, 画像ラボ, 2003. (掲載予定)

## 表彰

1. 電子情報通信学会, パターン認識・メディア理解研究会 第3回アルゴリズムコンテスト入賞. 1999年9月
2. 電子情報通信学会, 2000年度学術奨励賞. 2001年3月
3. 奈良先端科学技術大学院大学 情報科学研究科 2000年度最優秀学生賞. 2001年3月
4. 平成13年度電気関係学会関西支部連合大会奨励賞. 2002年4月

## 付録

### A. Tsaiの手法によるカメラの内部パラメータのキャリブレーション

本論文では、プログラムが公開され、一般的に広く用いられている Tsai のカメラキャリブレーション手法を用いてカメラの内部パラメータを推定する。本章では、まず Tsai の手法が用いるピンホールカメラモデルについて、三次元座標とその画像上への投影座標の関係式をまとめ、同手法によるカメラパラメータの推定手法について簡単に述べる。

#### A.1 ピンホールカメラモデル

一般に撮影に利用されるカメラは、複数のレンズを組み合わせることで CCD 上に画像を結像させるため、複雑な光学特性を持ち、その精密なモデル化は容易ではないといえる。しかし、レンズと投影面の距離に対して投影面から撮影対象までの距離が十分大きい場合、空間中の三次元座標と投影面上の幾何学的な対応を、ピンホールカメラモデルと呼ばれる投影モデルを用いて少数のパラメータで近似することが一般的に行なわれている。Tsai の手法においても、ピンホールカメラモデルを用いた投影モデルの近似を行なう。

ピンホールカメラモデルは、図 A.1 に示すように、投影面から距離  $F$  の位置に投影面と平行な面を置き、その面上に穴 (ピンホール)  $C$  をあけ、ピンホール  $C$  を通る光線のみが画像面上に結像するというモデルである。このカメラモデルでは、図 A.1 に示すようなピンホール  $C$  を原点とするカメラ座標系において、空間中の任意の三次元座標を  $s = (X, Y, Z)^T$ 、投影面上への投影座標を  $c = (u, v)$  とすれば、以下の関係が成り立つ。

$$(u, v) = \left(-F \frac{X}{Z}, -F \frac{Y}{Z}\right) \quad (50)$$

実際にはレンズを通過する光が屈折するため、三次元座標は  $s$  の投影座標には、投影中心に対する放射状の歪みが生じる。Tsai の手法では、レンズ歪み係数  $k_1, k_2$



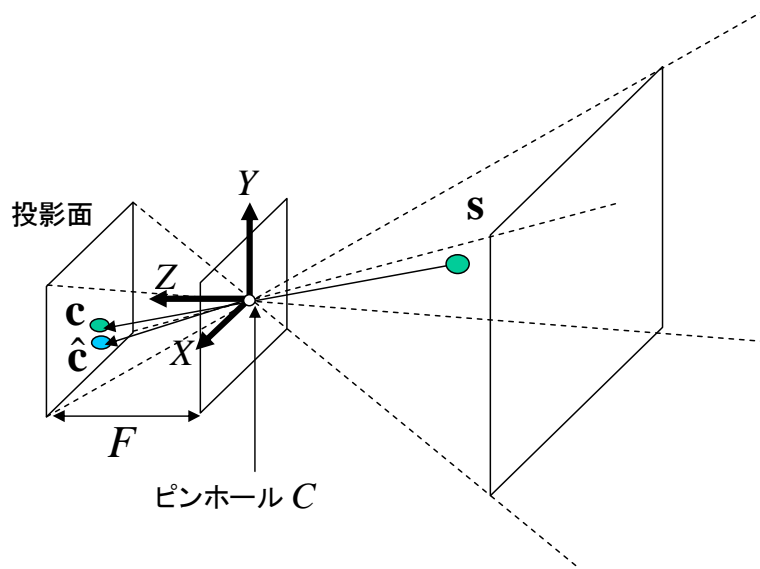


図 A.1 ピンホールカメラモデル

を用いてレンズ歪みを以下の式でモデル化している.

$$(u, v) = ((1 + k_1 r^2 + k_2 r^4)\hat{u}, (1 + k_1 r^2 + k_2 r^4)\hat{v}) \quad (51)$$

$$r = \sqrt{\hat{u}^2 + \hat{v}^2} \quad (52)$$

ただし,  $\hat{c} = (\hat{u}, \hat{v})$  は, レンズ歪みの影響によって実際に投影される画像面上の投影座標である.

## A.2 投影座標からデジタル画像への変換

投影面上に撮像された画像は、CCD の各素子によって離散化され、デジタル画像に変換される。カメラ座標系における CCD のサイズを  $(ccd_x, ccd_y)$ 、デジタル画像の画素単位での画像サイズを  $(scr_x, scr_y)$  とし、デジタル画像における投影中心の座標を  $(c_x, c_y)$  とすれば、投影座標  $(u, v)$  からデジタル画像上の座標  $(x, y)$  への変換は、以下の式で表される。

$$(x, y) = \left( \frac{scr_x}{ccd_x} s_x u + c_x, \frac{scr_y}{ccd_y} v + c_y \right) \quad (53)$$

ただし、 $s_x$  は CCD のサイズに対するスケール係数である。逆に、投影座標  $(u, v)$  からデジタル画像上の座標  $(x, y)$  への変換は、以下の式で表される。

$$(u, v) = \left( \frac{ccd_x}{scr_x s_x} (x - c_x), \frac{ccd_y}{scr_y} (y - c_y) \right) \quad (54)$$

この式を用いて、デジタル画像上で観測された任意の画素をカメラ座標系に変換し、次に式 (51), (52) によって歪みを取り除くことで、式 (50) で表される歪みのないピンホールカメラモデルのカメラ座標系に変換することが可能となる。また、式 (51) から (54) をまとめることで、以下の式を用いて、撮影したデジタル画像上における任意の座標  $(\hat{x}, \hat{y})$  を、歪みのないデジタル画像座標に補正することができる。

$$(x, y) = \left( (1 + k_1 r^2 + k_2 r^4)(\hat{x} - c_x) + c_x, (1 + k_1 r^2 + k_2 r^4)(\hat{y} - c_y) + c_y \right) \quad (55)$$

$$r = \sqrt{\left( \frac{ccd_x}{scr_x s_x} (\hat{x} - c_x) \right)^2 + \left( \frac{ccd_y}{scr_y} (\hat{y} - c_y) \right)^2} \quad (56)$$

### A.3 Tsaiによるカメラキャリブレーション手法

Tsaiによるカメラキャリブレーション手法についてまとめる。Tsaiの手法は、CCDのサイズ( $ccd_x, ccd_y$ )およびデジタル画像のサイズ( $scr_x, scr_y$ )を既知とし、多数のマーカの世界座標における三次元位置とその二次元座標の関係から、カメラの内部パラメータおよび外部パラメータを推定する手法である。実際にC言語のソースプログラムとして無料で配布されているため、コンピュータビジョンの研究者の間で広く利用されている。

まず、Tsaiのカメラキャリブレーション手法のアルゴリズムについて概要を述べる。Tsaiの手法によって推定されるカメラパラメータは、カメラの外部パラメータとして、回転行列(3パラメータ)とカメラの平行移動ベクトル(3パラメータ)、カメラの内部パラメータとして、焦点距離 $F$ 、レンズ歪み係数 $k_1, k_2$ 、スケール係数 $s_x$ 、画像中心座標( $c_x, c_y$ )の合計12パラメータである。同手法では、まず12のパラメータそれぞれに対して、線形演算による近似値の算出を行なう。この線形演算によるパラメータの算出方法は、パラメータ毎に異なるため、ここでは個々の手法についての説明は省略する。次に、得られた近似値を用い、非線形最適化の手法を用いて画像上の投影誤差を最小化することで、各パラメータの最適化を行なう。

次に、手法の入力として使用するマーカの座標の検出について述べる。キャリブレーションに用いるマーカとしては、一般的に平面や立方体といった幾何学的な形状を容易に測定できる物体上に作成した格子パターンの格子点が利用される。このようなマーカをカメラで撮影し、画像上で格子点の座標を検出することによって、多数のマーカの二次元座標と三次元座標を容易に得ることができる。ただし、Tsaiの手法においては、多数のパラメータを同時に推定するため、精度の良いパラメータを得るためには少なくとも数百点のマーカを用いることが必要である。

## B. GUIツールによるボクセルモデルの修正

本論文で述べた、受動的な動画像からの三次元モデル化においては、鏡面性の物体やテクスチャの無いビル側面などでモデルの欠損がおこる。そこで、GUIベースでのボクセル修正ツールを試作し、ボクセルデータの修正を試みた。以下では、ボクセル修正ツールの主な機能と、本文で示した町並みの復元結果の修正例について述べる。

本ツールは、図B.1(1)に示すツールバーと、(2)に示すモデルビューワからなる。ツールバーにはそれぞれ機能が割り当てられており、ツールバーのアイコンを選択した状態で、モデルビューワ部分での操作を行うことで、モデルの修正が可能となる。モデルの修正操作は、図B.2に示すように、ペイントツール感覚の二次元操作で、修正箇所を塗りつぶすことによって行われる。また、モデルビューワの視点位置および視線方向は、モデルビューワ上でマウスのドラッグを行うことで可能である。以下に、本ツールが持つ機能一覧を示す。

- (1) モデルの読み込み ボクセルファイルの読み込みを行う。
- (2) モデルの書き出し 修正済みボクセルファイルを書き出す。
- (3) 表示モードの変更 モデル表示を奥行き画像表示モードに変更する。
- (4) 消しゴムツール モデル上のノイズ・エラーの削除を行う。
- (5) 面追加ツール モデルの穴部分を面で埋め、テクスチャを貼り付ける。
- (6) テクスチャ変更ツール モデルは変更せずに、テクスチャのみ更新する。
- (7) 平面化ツール モデルを平面化し、新たなテクスチャを貼り付ける。
- (8) テクスチャ選択ボタン テクスチャとして貼り付ける動画像系列を選択する。

図B.3に、本ツールを用いて修正した三次元モデルを示す。モデルの修正では、空に見られた復元エラーの削除、ビル側面の平面化、穴の修正などの操作を筆者が10分程度行った。同図から、修正前に見られたモデルの穴が修正され、ビルの側面形状が復元されている。

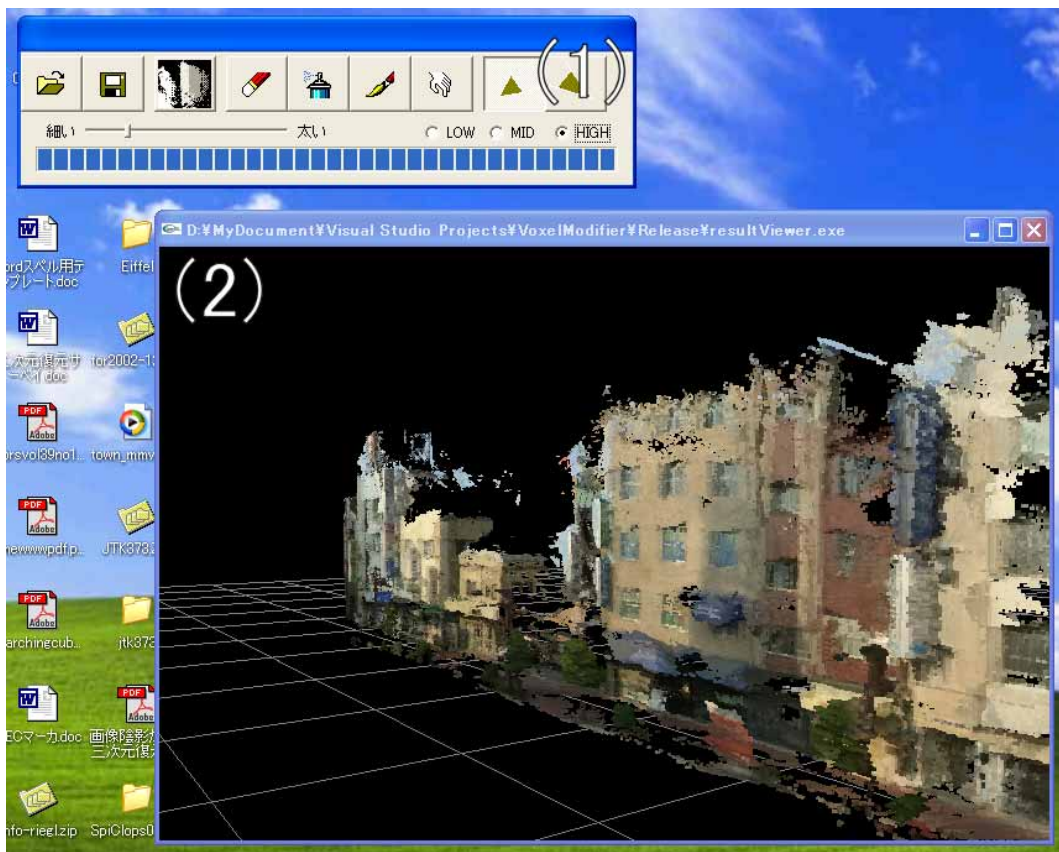
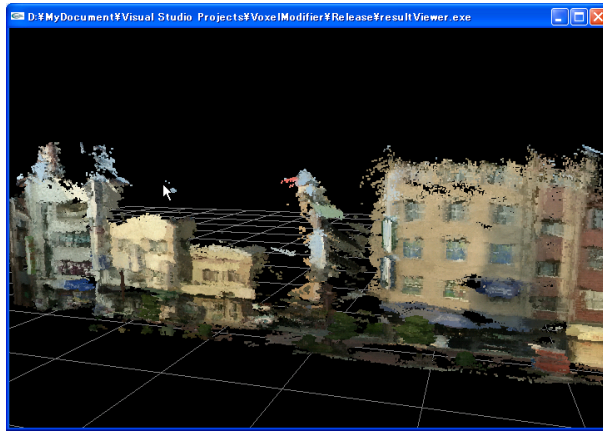
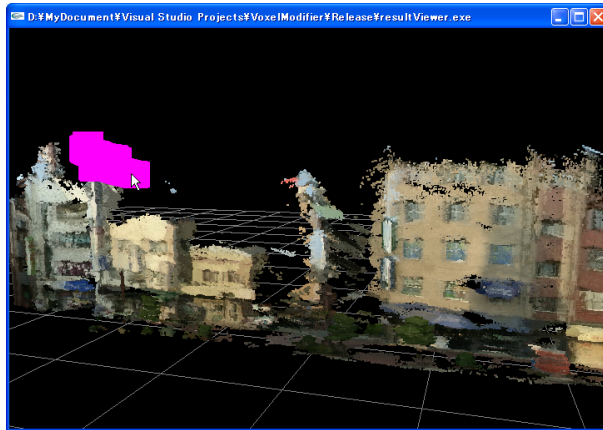


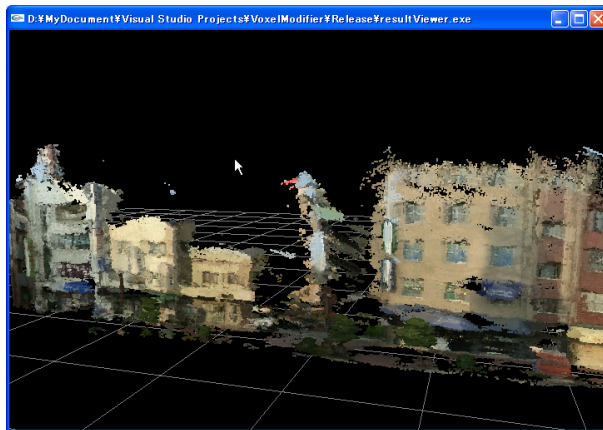
図 B.1 ボクセル修正ツールの構成



(a) 修正前



(b) ツール使用中



(c) 修正後

図 B.2 消しゴムツールの使用例

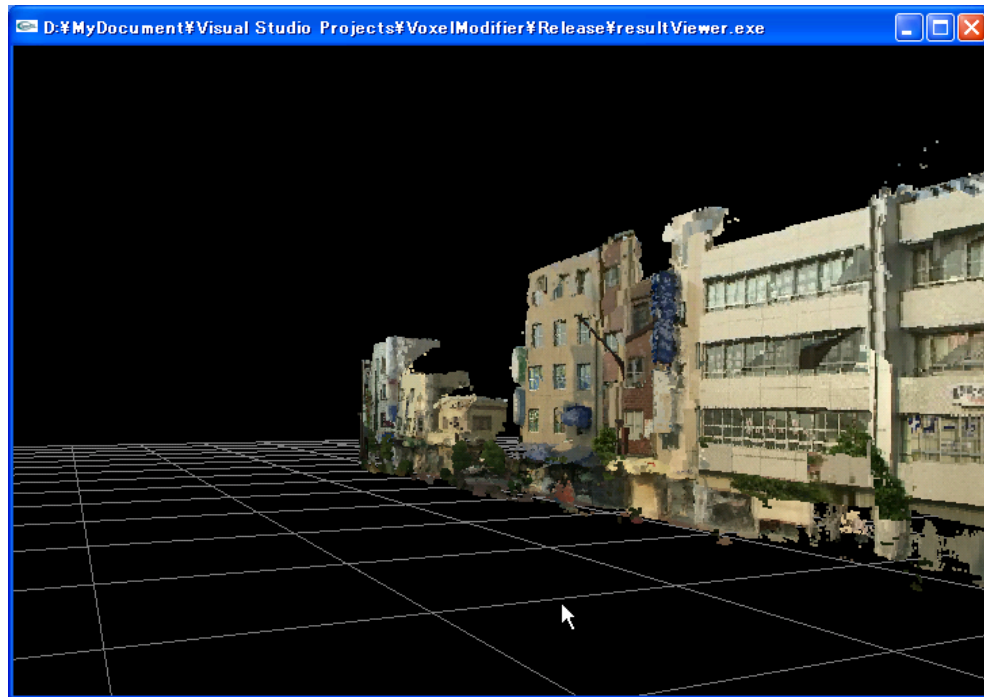
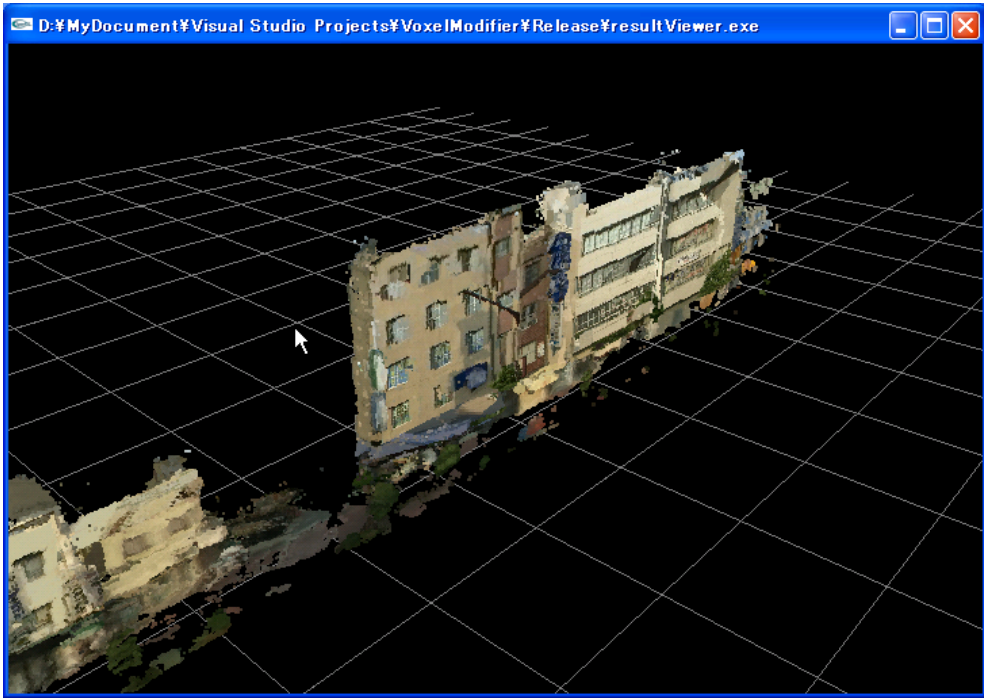


図 B.3 ツールによる修正後のモデル