

NAIST-IS-MT9751115

修士論文

多視点ステレオ画像を用いた屋外環境構造の 再構成に関する研究

横光 澄男

1999年2月12日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

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

横光 澄男

指導教官： 横矢 直和 教授
千原 國宏 教授
竹村 治雄 助教授

多視点ステレオ画像を用いた屋外環境構造の 再構成に関する研究*

横光 澄男

内容梗概

屋外環境構造を計算機内に再構成することは、都市工学や建築工学など広範囲な応用が期待される重要な研究課題である。しかし、そのためのデータ獲得及びモデル構築は、従来、手作業に依存している部分が多く、コストが高いため容易にモデルを作成することは困難である。このような理由から、効率的、自動的なモデル構築技術に対する要求が高まっている。

そこで本論文では、屋外環境構造を計算機内に再構成するために、多視点のステレオ画像から仮想空間を構築する手法を提案する。屋外環境下のステレオ撮影に関して、2つのカメラの光軸が互いに平行で、かつ各々の投影面が同一平面上に配置される標準ステレオ撮像系を構成することは困難であると考え、提案手法では複数視点からの撮影時に汎地球測位システムと3軸角度センサを用いて、各撮影地点間の位置、姿勢の関係を得る。この関係からエピポーラ線の平行化を行ない、ステレオ法を用いて各視点での奥行き情報を獲得する。得られた各視点での奥行き情報を用いて作成する3次元モデルを統合して、屋外環境構造の再構成を行う。実際に、各センサを用いて撮影したステレオ画像から、屋外環境の3次元モデルを再構成し、本手法の有効性を確認した。

キーワード

仮想環境, 屋外環境構造再構成, 多視点ステレオ画像, エピポーラ線平行化

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9751115, 1999年2月12日.

Reconstruction of Real Outdoor Environmental Structures Using Multiple Stereo Images*

Sumio Yokomitsu

Abstract

This paper describes a technique to reconstruct structures of outdoor environments in a virtual space from multiple stereo images of a real scene. Reconstruction of an outdoor environment has many potential applications in many areas, such as civil engineering and architectural engineering. Data acquisition and generation of models require much computational cost, and many parts of the tasks depend on manual operation. Therefore, it is desired to reconstruct the model efficiently and automatically.

Because it is difficult to compose a simple calibrated camera system for stereophotography with known external camera parameters in outdoor environments, the proposed method acquires the geometric relation between two cameras using a global positioning system and a gyro sensor. Rectification of images is conducted using this relation and 3-D information is obtained from rectification images. Then, a realistic virtual space is generated by integrating 3-D information obtained from multiple viewpoints. Finally, images of the constructed virtual space from an arbitrary viewpoint is shown to confirm the precise reconstruction of outdoor scene.

Keywords:

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9751115, February 12, 1999.

Virtual Environment, Reconstructed Outdoor Environment, Multiple Stereo Images, Rectification

目次

1. はじめに	1
2. 関連研究：実環境の表現手法	3
2.1 イメージベースレンダリング	3
2.2 ポリゴンベースレンダリング	4
2.3 本研究の位置付けと方針	5
3. 屋外環境構造の再構成	7
3.1 ステレオ視による奥行き情報の獲得	7
3.1.1 ステレオ視の原理	9
3.1.2 エピポーラ線の平行化	12
3.1.3 ステレオ画像の対応付け	17
3.2 奥行き情報からの平面の推定	22
3.2.1 画像の領域分割	24
3.2.2 平面の推定	28
3.3 多視点データの統合	32
4. 実験	35
4.1 実験環境	35
4.2 実験結果と考察	39
5. 考察	48
6. おわりに	50
謝辞	51
参考文献	52
付録	I

A. カメラモデルと射影行列	I
A.1 透視投影モデル	I
A.2 カメラパラメータ	III
A.2.1 カメラ外部パラメータ	III
A.2.2 カメラ内部パラメータ	IV
A.3 同次座標系表示	V
B. ロバスト推定	VI

目 次

1	屋外環境構造再構成のための処理の流れ	8
2	標準ステレオ撮像系	11
3	一般的な撮像系	11
4	エピポーラ線平行化の概念図	13
5	エピポーラ線平行化のための再投影	13
6	カメラ座標系間の関係	14
7	HSV6 角錐カラーモデル	21
8	平面推定における処理の流れ図	23
9	領域分割処理手順	23
10	多数決フィルタの例	27
11	クラスタ間統合説明図	27
12	ロバスト推定	29
13	平面推定アルゴリズムの説明図	31
14	カメラ座標系と世界座標系間の関係	32
15	GPS の外観	36
16	3 軸角度センサの外観	36
17	実験におけるカメラ配置	37
18	エピポーラ線平行化処理結果 (ステレオペア 2)	42
19	視差推定処理結果 (ステレオペア 2)	43
20	領域分割結果 (図 18(a) の基準画像)	44
21	ステレオペア 2 による 3 次元モデルの獲得	45
22	ステレオペア 1 による 3 次元モデルの獲得	45
23	ステレオペア 3 による 3 次元モデルの獲得	46
24	ステレオペア 4 による 3 次元モデルの獲得	46
25	統合後の 3 次元モデル	47
26	ピンホールカメラモデル	II
27	カメラ座標系と透視変換	II
28	実際のカメラモデル	III

表 目 次

1	各センサの実測データ	37
2	実験において利用したパラメータ	38

1. はじめに

近年，計算機の処理能力や，グラフィックスハードウェアの飛躍的向上を背景にして，計算機により構築された仮想環境を利用するシステムの研究が盛んに行なわれるようになってきた．この仮想環境を用いて，科学的可視化，マンマシンインターフェースや，建築，医学，芸術などの分野で多数の研究が報告されている [1, 2, 3, 4]．現実世界の環境を計算機内に仮想環境として再構成することで，現実では不可能な視点からの観察や，環境アセスメントに関するさまざまなシミュレーションなどを行なうことが可能となり，今までにない新しい応用技術が実現されると期待されている．

しかし，現実環境をコンピュータグラフィックスのみで再構成する場合，再構成された仮想空間の写実性や臨場感には限界がある．また，正確な現実環境を再構成しようとする場合，人間の手による多大な作業を必要とする場合が多い．再構成される仮想空間の写実性や臨場感を高め，また人手による作業を減らすために，従来よりさまざまな研究が提案されている [5, 6, 7, 8, 9, 10]．特に再構成する対象物体が屋外環境下に存在する場合，その撮影環境や照明条件の複雑さから，ほとんどの再構成手法が手作業に依存している [10, 11, 12, 13]．そこで本論文では，屋外環境下に存在する対象物体の3次元モデルを自動で，かつ正確に獲得することを目的とし，これを実現するために，各視点における屋外の実写画像をもとに再構成する手法を提案する．

現実世界の環境を仮想空間内に再構成する手法は大きく2つに分類することができる．1つ目は，現実世界の情景をさまざまな視点からの実写画像としてあらかじめ蓄積しておき，観察時には視点の移動に応じた画像を検索して提示する，イメージベースレンダリングである [14, 15, 16]．この手法は撮影された2次元画像をそのまま提示しているためリアリティは高い．しかし，この手法では撮影された範囲でしか視点を動かすことができず，視点移動の自由度を高めるためには膨大な量の画像が必要となる．このため蓄積するデータ量が大きい，提示画像の検索に時間がかかるなどの問題が生じる．この手法を改良し，複数枚の画像から，その間の視点位置画像を補間処理で求めることにより，撮影する画像数を削減し視点操作の自由度を高める研究も行なわれている [17]．しかしこの手法にお

ける補間処理は、人間が2次元画像から得る知識を用いて手動で行なっているため、前処理に多大な時間がかかるという問題が生じる。

2つ目は、現実世界の環境の3次元モデルを獲得して仮想空間内に再構築するポリゴンベースレンダリングである [11, 18, 19]。対象物体の3次元モデルを完全に獲得できれば、任意の視点位置からの観察が可能な仮想物体として扱うことができ、蓄積するデータ量も小さいという特徴がある。しかし、正確な3次元モデルを獲得するには莫大な計算コストが必要であり、十分な信頼性を得るためには手作業による補助を必要とする手法が多い。そこで、再構成される3次元モデルの信頼性を高めるため、多くの入力画像を用いることで、現実環境に関する重複した情報を利用したり、特徴領域や前提知識を基に現実環境を計算機に理解させ、処理の自動化を計る研究も行なわれている [20]。

イメージベースレンダリングの手法は基本的に3次元モデルを持たないため、現実世界の環境をシミュレーションする用途には不向きであると考えられる。そこで本論文では、一旦3次元モデルを構築すれば、その後の応用範囲が広いポリゴンベースレンダリングの手法に注目し、屋外の広範囲な撮影環境に対応するために、多視点のステレオ実画像から仮想空間を自動で再構成する手法を提案する。提案手法では、複数視点からの撮影時に汎地球測位システム (Global Positioning System ; 以下 GPS) と3軸角度センサを用いて、各撮影位置間の位置・姿勢の関係を得る。この関係からエピポーラ線の平行化を行ない、ステレオ法を用いて各々の視点における奥行き情報を獲得する。得られた各々の視点での奥行き情報を用いて作成する3次元モデルを統合して、屋外環境構造の再構成を行なう。

以下2章では、従来の実環境の表現手法を概観し、本研究の位置付けと方針について述べる。3章では本研究で提案する屋外環境構造の再構成手法について詳細に述べる。4章で、実際に撮影した画像から、提案手法を用いて実験を行ない、その実験結果を示す。最後に提案手法の考察を述べ、まとめと今後の課題について述べる。

2. 関連研究：実環境の表現手法

現実世界の環境を計算機内に再構成する手法については，建築やアミューズメントの分野などそのアプリケーションの多様性から，数多くの研究が報告されている [5, 12, 13, 15]. 本章では，まず，現実世界の環境を計算機内に再構成する従来手法を概観する．2.1節でイメージベースレンダリングについて，2.2節でポリゴンベースレンダリングについて述べ，現状の問題点をそれぞれ整理する．最後に 2.3節で本研究の位置づけと方針について述べる．

2.1 イメージベースレンダリング

イメージベースレンダリングはあらかじめ現実世界の情景をさまざまな視点からの実写画像として蓄積しておき，視点の移動に応じた画像を検索し，提示するという手法である．このイメージベースレンダリングの代表的なものとして，Aspen Moviemap[14] が挙げられる．この手法は撮影位置が既知である 2 次元入力画像を，ユーザの視点移動に応じて提示する手法である．また，別の手法として QuickTimeVR[15] が挙げられる．QuickTimeVR はあらかじめ計算機内にパノラマ画像を持っており，その中から利用者の任意視点の画像を提示するシステムである．しかしながらこれらの手法は，撮影された視点位置からしか空間内を見渡せず，仮想環境内を自由に視点移動することができない．これらの手法に代表されるように，イメージベースレンダリングでは再構成された仮想環境内において，視点移動に対する制約が厳しく，視点移動の自由度を高めるためには膨大な量の画像が必要となる．そのため，画像データ量を削減するための工夫として，さまざまな視点の画像から代表的な画像を選択し，変形予測と輝度予測誤差を用いて膨大なデータを効率的に記述し蓄積するアプローチも提案されている [21].

大規模な仮想都市空間の再構成を目的としたイメージベースレンダリングの手法としては，遠藤らの手法が挙げられる [12, 13]. この手法の特徴は，見回し方向の自由度を確保するため，水平画角 60 度のビデオカメラを全周方向に 7 台配置した車両搭載型の撮影システムを構築しており，この 7 台のカメラにより撮影された画像からパノラマ画像を生成し，提示画像としている．生成されたパノ

ラマ画像により並進 2 自由度, 見回し 1 自由度の合計 3 自由度の視点移動を可能にしている. この手法は各パノラマ画像において画像のデータ圧縮等を行なっておらず, データ量が非常に大きくなるという欠点がある. また, 膨大なデータ量のため, 画像提示速度も秒間数フレームとなっており, 提示画像の不連続表示も大きな問題の一つである.

2.2 ポリゴンベーストレンダリング

ポリゴンベーストレンダリングは現実世界の環境の 3 次元形状を獲得して仮想空間内に再構成する手法である. この手法の最大の利点は, 対象物体の 3 次元形状を完全に獲得できれば, 視点の位置変化や衝突判定などが可能な仮想物体として扱うことができ, 蓄積するデータ量も小さいという点である. このような理由により, 従来から現実環境の 3 次元形状を獲得するために, さまざまな手法が提案されている.

対象の 3 次元形状を得る代表的な手法の一つとして, 特徴点追跡による因子分解法が挙げられる [9]. この手法は移動カメラで取得した時系列画像上の特徴点の対応を取り, その各画像上の対応点の位置を成分に持つ計測行列 \mathbf{W} を, カメラまたは対象の運動を表現する行列 \mathbf{M} と対象の形状を表現する行列 \mathbf{S} の積に $\mathbf{W} = \mathbf{MS}$ と分解することで, 時系列画像から対象形状と各時系列画像間の相対位置を同時に求める手法である. この因子分解法が安定して 3 次元形状を獲得できる最も大きな要因は, 画像の投影モデルとして線形な正射影モデルを用いたことにある. この仮定のため, 正射影近似が当てはまらない条件下での 3 次元形状復元の場合, つまり対象物体の奥行きが大きく変化するような場合は計算結果に大きな誤差が生じてしまう結果となる. この線形な投影モデルでの因子分解法を非線形な透視投影モデルに拡張した研究 [19] も行なわれているが, 前提条件や特徴点の手動での対応付けなど残された問題は多い.

もう一つの代表的な手法として, ステレオ視による 3 次元モデルの獲得が挙げられる. ステレオ視は 2 台のカメラにより撮影された画像を用いて, 3 角測量の原理により環境の奥行きを受動的に計測する手法である. この方法は, 設備が簡単で, 適用できる環境に対する制限が少なく, 更に広範囲の奥行き情報を非接触,

かつ直接的に獲得することができる。そして、カメラモデルとして透視投影モデルを扱うことが可能である。しかし、ステレオ視を用いる場合の問題点として、左右のステレオ画像に撮影されている点同士の対応を求める対応点探索問題の難しさや、厳密にキャリブレーションされたステレオ撮像系の構築が難しいことが挙げられる。

特に対象物体が屋外環境下の場合、その撮影環境や照明条件の複雑さからほとんどの手法が、3次元情報の獲得及び3次元形状の再構成に関して、処理の大部分を手作業に依存している。さらに、再構成する対象物体が大規模になると、これらの手作業には莫大な時間と労力が必要になり、自動的、かつ効率的な屋外環境構造の再構成手法の確立が望まれる。

2.3 本研究の位置付けと方針

屋外環境構造の対象物を再構成するためには、再構成されたデータ量が少ないこと、画像提示時の視点移動にともなう時間遅れが少ないこと、任意の視点移動が可能であることなどが挙げられる。そのため、イメージベーストレンダリングを用いた場合、再構成する対象物体が大規模になるほど、蓄積するデータ量は多く、データ量に伴う提示画像の時間遅れも問題となり、視点移動にもかなりの制約が生じる。そこで本論文ではポリゴンベーストレンダリングの利点を生かした屋外環境構造の再構成を行なう。また、対象物体の3次元モデルを構築するための奥行き計測手法としては、屋外環境の複雑な照明条件でも、計測が行なえなければならない。また、撮影場所に制約を受けない撮影方法も必要となる。そこで、本論文では、この計測方法としてステレオ法を採用する。

しかし、屋外環境下でステレオ視により奥行き情報を得ようとする場合、2つの問題点が挙げられる。その問題とは、照明条件や撮影条件の厳しい屋外環境下において、厳密な標準ステレオ撮像が行なえないということ、そして、ステレオ視において計測される奥行き情報には、不正確な値も含まれている可能性が高いということである。一つめの問題に対しては、対象物体の撮影時にGPSと3軸角度センサを用いて、各撮影地点間の位置・姿勢の関係を得ることで解決する。つまり、各センサにより得られた撮影地点の位置・姿勢の関係から、視点の異な

るステレオ画像間に成立する幾何学的関係に基づいてエピポーラ線の平行化を行なう。この処理により一般的な撮像系で撮影されたステレオ画像を、標準ステレオ撮像系の画像に変換する。2つ目の問題に対しては、ステレオ視により得られる奥行き情報に対して、不正確な奥行き情報に影響を受けにくい3次元モデル構築手法を提案する。具体的には、基準ステレオ画像に領域分割を施し、分割された領域はそれぞれ一つの平面を構成すると仮定し、得られた奥行き情報から領域毎の平面方程式をロバスト推定を用いて推定する。これにより各視点における3次元モデルを獲得し、各カメラの位置・姿勢情報を基に統合し、屋外環境構造の再構成を行なう。

3. 屋外環境構造の再構成

本章では，屋外環境構造を再構成するための提案手法について述べる．提案手法の処理は以下の3段階の手順で行なわれる．各処理手順を図1に示し，順に説明する．

1. ステレオ視による奥行き情報の獲得 (図1(a))

標準ステレオ撮像系での撮影が困難な屋外環境下において，2つのカメラ間の位置・姿勢の関係からエピポラ線を平行化し対応点探索を行なう．この処理により得られた視差マップから奥行き情報を獲得する．

2. 奥行き情報からの平面の推定 (図1(b))

入力画像に領域分割を施し，分割された領域はそれぞれ一つの平面を構成すると仮定し，得られた奥行き情報から領域毎の平面方程式をロバスト推定を用いて決定する．

3. 多視点データの統合 (図1(c))

各視点から得られた3次元モデルを，各撮影カメラ間の位置，姿勢情報を基に統合する．

以降では，各段階での処理について順に詳細を述べる．

3.1 ステレオ視による奥行き情報の獲得

本節では，入力されるステレオ画像に対し，奥行き情報を獲得するための処理について述べる．以降，3.1.1節ではステレオ視の原理について，3.1.2節ではステレオ視において重要な拘束条件であるエピポラ拘束について説明し，エピポラ線を平行化する手法について述べる．そして，3.1.3節で，ステレオ視の対応付け手法について述べる．

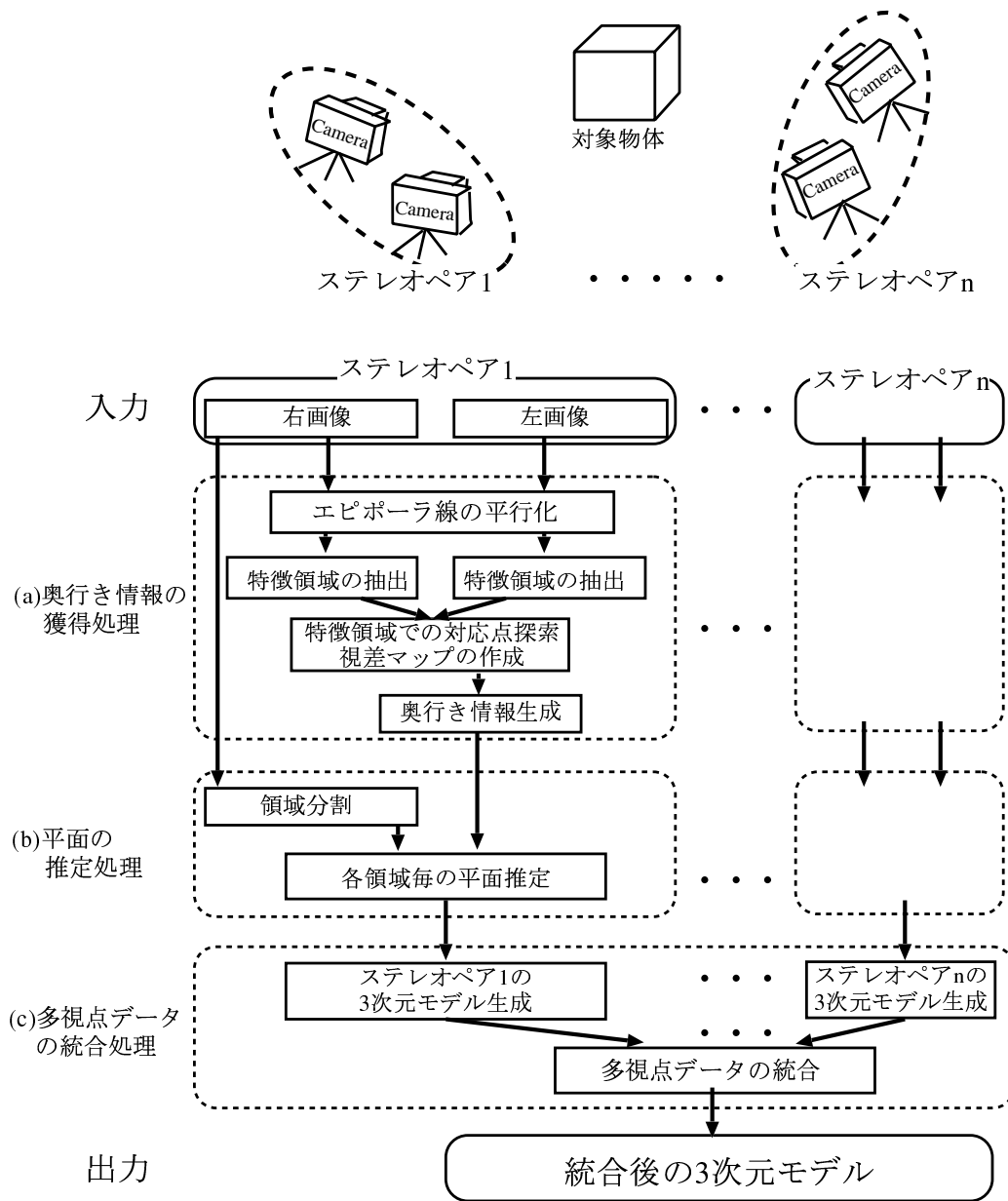


図 1 屋外環境構造再構成のための処理の流れ

3.1.1 ステレオ視の原理

ステレオ視はカメラなどの視覚センサを異なる位置に配置し、異なる視点から同一の対象を観測し、それぞれの画像上への投影位置の違いから、対象の奥行き情報を得る手法である。その原理は3角測量法に基づいており、我々人間も同様な原理により、現実の世界の奥行き情報や3次元構造の把握のための重要な情報を得ている。3角測量法はあるがままの世界をそのまま観測する受動的な計測法の代表であり、古くからもっとも一般的な距離測定法として用いられている。

ステレオ視のもっとも標準的な構成は、図2に示すような標準ステレオ撮像系と呼ばれるものである。この構成は、焦点距離が等しいカメラ2台に関して、光軸が互いに平行で、かつ各々の画像面が同一平面上に配置される場合に成立する。

図2に示すように空間中の点 P の座標系を、2つのレンズ中心を結ぶ線分(基線)の2等分点を原点として定め、 X 軸を基線と平行な方向、 Y 軸を基線と光軸に垂直な方向、 Z 軸を光軸方向とした座標系で表す。また、各画像座標系を各画像面の中心を原点と定め、水平方向に x 軸、垂直方向に y 軸方向とした座標系で表す。点 P の画像平面上の像がそれぞれの画像座標系の $P_l(x_l, y_l), P_r(x_r, y_r)$ に投影されたとすると、点 P の座標値 (X, Y, Z) には以下の関係式が成立する。

$$\begin{aligned} X &= b \cdot \frac{(x_l + x_r)/2}{d} \\ Y &= b \cdot \frac{(y_l + y_r)/2}{d} \\ Z &= b \cdot \frac{f}{d} \end{aligned} \quad (1)$$

ただし f は焦点距離、 b は基線長、 d は視差で、

$$d = x_l - x_r \quad (2)$$

である。したがって、 f, b が既知のもとで、左右の画像上の投影座標の対応関係を表す視差から、もとの3次元位置 $P(X, Y, Z)$ が計算される。さらに、投影座標の対応関係に関して、標準ステレオ撮像系におけるカメラ配置では常に

$$y_l = y_r \quad (3)$$

という関係が成立する。つまり、左右の投影点の y 座標は必ず等しくなるので、対応点を見つける際、同一の水平ライン上を探せばよいことになる。

ここで、ステレオ視を用いて屋外環境構造の再構成を行なおうとする場合、この標準ステレオ撮像系を仮定すると幾つかの問題が生じる。なぜなら屋外環境構造のような室内の対象物に比べ非常に大きな対象物を計測する場合、基線長を十分長く取る必要があるが、基線長を長くとり、かつ厳密にキャリブレーションが行なわれた標準ステレオ撮像系を屋外で構成するという仮定は、現実的ではないからである。

このため、標準ステレオ撮像系を仮定せずに、一般的な撮像系を用いる。この一般的な撮像系の場合(図3)においても、一方の画像上(基準画像)の任意の点に対する他方の画像上(参照画像)の対応点は、基準画像の投影中心と投影点、および参照画像の投影中心の3点で決定される平面と参照画像の投影面の交線上に存在する。この直線のことをエピポーラ線と呼び、ステレオ画像の対応点が満たすべきこのような拘束条件のことをエピポーラ拘束という。すなわち、撮影位置が既知である撮像系において、同じ3次元世界を複数の画像平面に投影すると、それらの画像間には必ずエピポーラ拘束が成り立ち、かつその拘束条件こそが画像間に存在する唯一の幾何学的関係である[22]。このエピポーラ拘束はステレオ視において非常に重要な意味を持つ。なぜなら、対応点探索問題においてエピポーラ線上での1次元の探索と画像平面上での2次元の探索では、その困難さと計算コストに大きな違いがあるからである。

先に述べたように、標準ステレオ撮像系の場合、ある点のステレオ画像上の投影点の y 座標は必ず等しくなる。つまり、両画像中のエピポーラ線は必ず同一 y 座標上(水平線上)になる。しかし、一般的な撮像系の場合、光軸の向きや焦点距離の違いなどから、必ずしもエピポーラ線が同一水平線上になるとは限らない。そこで次節では、一般的な撮像系で撮影された画像間に成り立つ幾何学的関係から、画像上のエピポーラ線の対応関係を求め、両画像中のエピポーラ線を同一水平線上に平行化する方法を示す[23]。

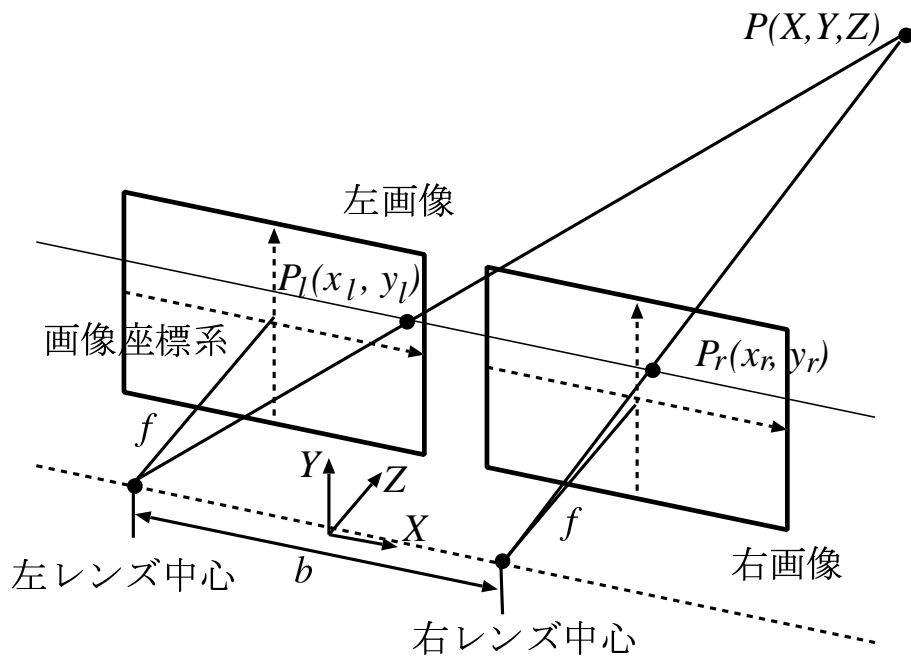


図 2 標準ステレオ撮像系

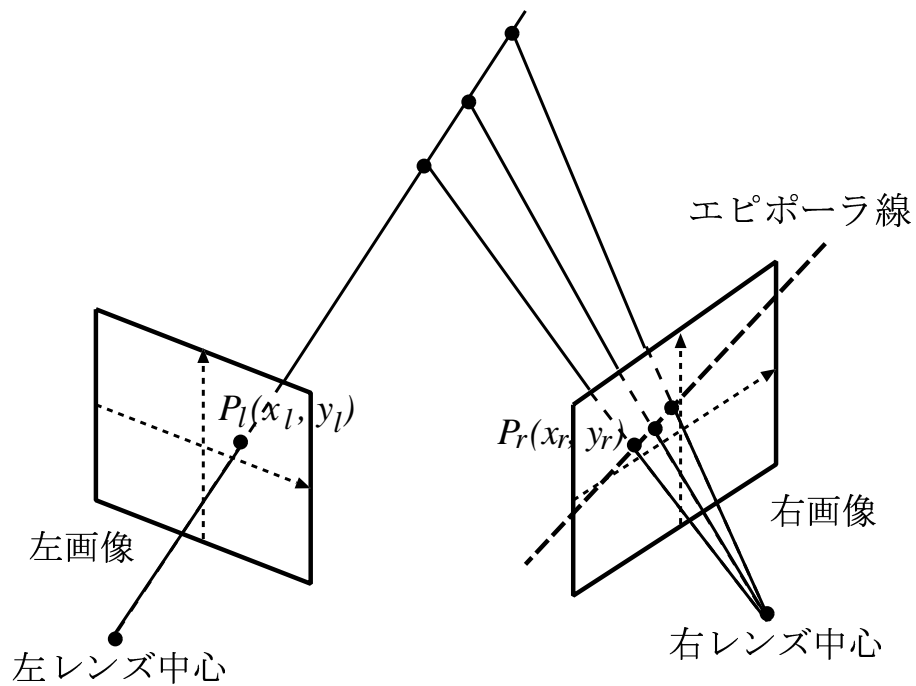


図 3 一般的な撮像系

3.1.2 エピポーラ線の平行化

エピポーラ線を平行化することにより，ステレオ視における対応点探索範囲は両画像中の，同一水平線上のみを探索すればよいことになる．この幾何的な意味は図4を用いて説明できる．それぞれのカメラの結像の特性(カメラ内部パラメータ)と2つのカメラの位置姿勢(カメラ外部パラメータ)¹が分かっているならば，対応点を探索するにあたりエピポーラ拘束を利用でき，この拘束条件をもとにエピポーラ線の平行化を行なう．

エピポーラ線の平行化の基本的な処理の概要は，図5のように2枚のステレオ画像 I_1, I_2 を，この2つの画像平面の交線 L の方向ベクトルと各カメラの光学中心を結ぶ線分の方向ベクトルに垂直なベクトルを法線ベクトルに持つ面 I に投影しなおすことである．この平面 I に投影するカメラの内部パラメータを同じ値とすることでカメラの光軸が平行で，焦点距離の等しいステレオ画像を得ることができる．

実際の変換処理は，図5に示すように，空間中のある点 M の各画像内の投影点 m_1, m_2 を新しい投影面 I 上の点 m'_1, m'_2 に投影しなおすことである．画像平面 I_1, I_2 を平面 I に投影する変換行列を求める処理の流れとしては，最初に画像平面 I_1, I_2 を投影しなおす平面 I の座標系 x, y, z 軸の各方向ベクトルを求め，その後，画像平面 I_1, I_2 を平面 I に投影しなおす変換行列を求める．以下，この処理手順を順を追って説明する．

図6に示すように，一般的な撮像系で構成されたステレオ撮像系を考える．このとき空間の点の位置を表す世界座標 (X, Y, Z) とその点の画像平面上の画像座標 (u, v) の関係は，それぞれのカメラの透視投影行列 P_1, P_2 (それぞれ 3×4) を用いた射影方程式として

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = P_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (i = 1, 2) \quad (4)$$

¹各カメラパラメータの詳細な説明は付録Aで述べる

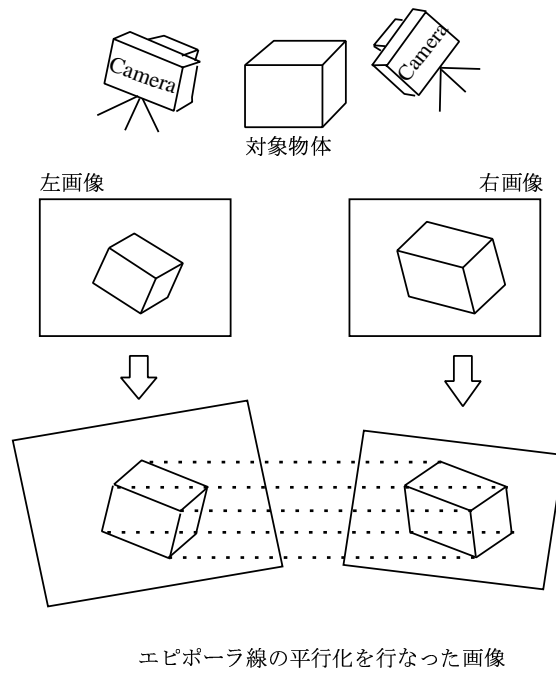


図 4 エピポーラ線平行化の概念図

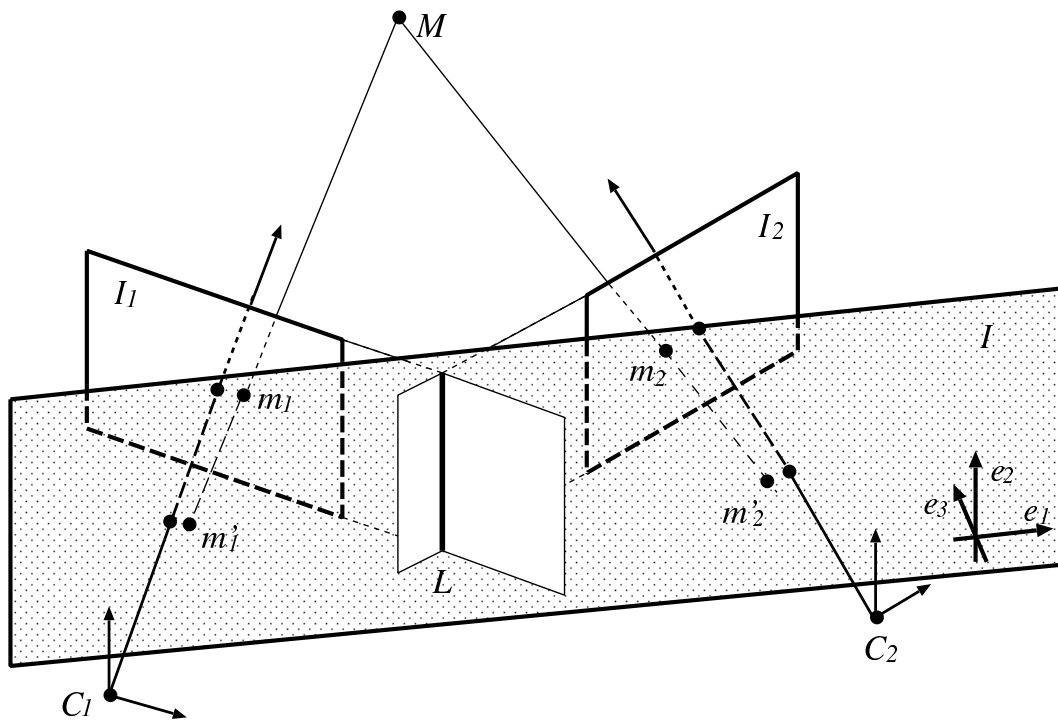


図 5 エピポーラ線平行化のための再投影

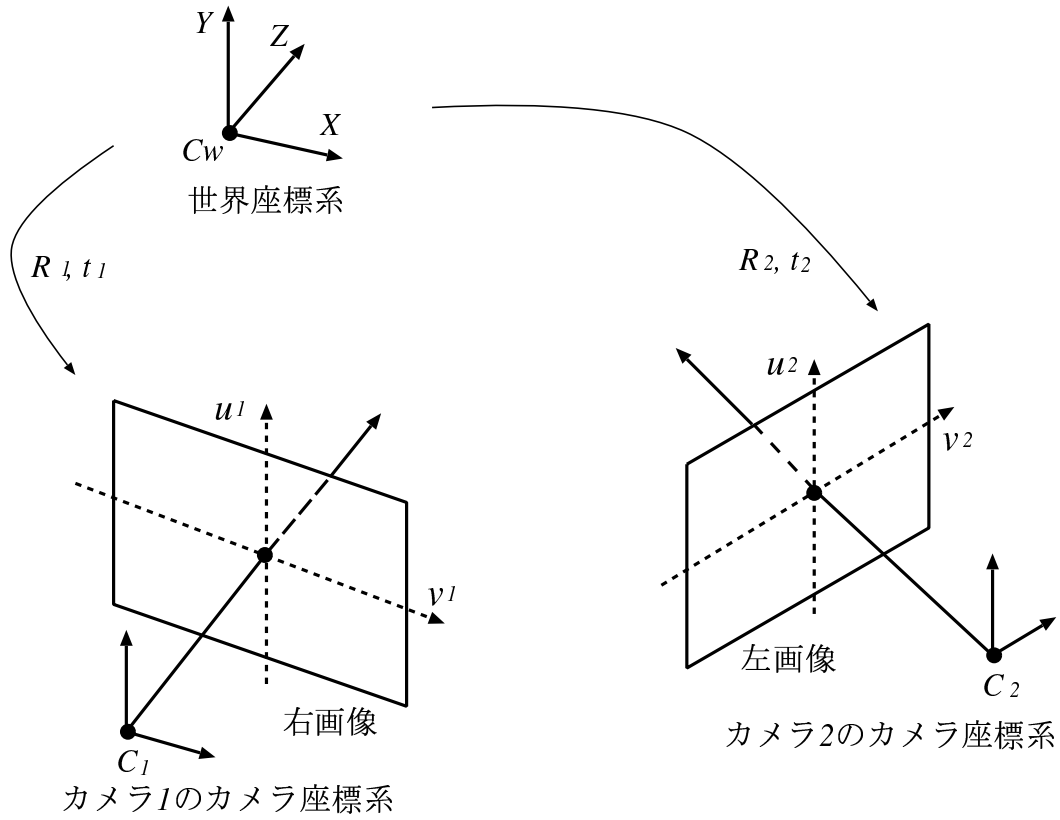


図 6 カメラ座標系間の関係

と記述される. また, 各カメラの内部パラメータ行列をそれぞれ \mathbf{A}_1 , \mathbf{A}_2 とし, 各カメラ座標系の世界座標系に対する相対的な位置, 姿勢を, カメラ 1, カメラ 2 のカメラ座標系に対し, それぞれ $\mathbf{R}_1, \mathbf{t}_1$, 及び $\mathbf{R}_2, \mathbf{t}_2$ で表すと, 各透視投影行列 $\mathbf{P}_i (i = 1, 2)$ は

$$\mathbf{P}_i = \mathbf{A}_i [\mathbf{R}_i \mid -\mathbf{R}_i \mathbf{t}_i] \quad (5)$$

と記述される. また, 図 5 に示してあるように, 空間の点の世界座標における位置を $\mathbf{M} = (X, Y, Z, 1)$, その点のそれぞれのカメラ画像面上での位置を $\mathbf{m}_1 = (u_1, v_1, 1)^T$, $\mathbf{m}_2 = (u_2, v_2, 1)^T$ とおいて, 式 (5) を用いて式 (4) を書き直すと,

$$\mathbf{m}_i = \mathbf{A}_i [\mathbf{R}_i \mid -\mathbf{R}_i \mathbf{t}_i] \mathbf{M}$$

$$= [\mathbf{R}'_i \mid \mathbf{t}'_i] \mathbf{M} \quad (i = 1, 2) \quad (6)$$

但し，射影行列 P は，

$$\mathbf{P}_i = [\mathbf{R}'_i \mid \mathbf{t}'_i] = \left[\begin{array}{c|c} \mathbf{r}'_{xi}{}^T & t_{xi} \\ \mathbf{r}'_{yi}{}^T & t_{yi} \\ \mathbf{r}'_{zi}{}^T & t_{zi} \end{array} \right] \quad (i = 1, 2) \quad (7)$$

となる．この射影行列の各 $\mathbf{R}'_i (3 \times 3)$, $\mathbf{t}'_i (3 \times 1)$ は， \mathbf{R}'_i がワールド座標系を基準とした，カメラ座標系の x, y, z 軸の方向ベクトルであり， \mathbf{t}'_i は世界座標系の原点とカメラ座標系の原点との並進ベクトルである．この式 (6) が，世界座標系と画像上の点との対応関係を表す射影方程式である．以下ここで導出した各カメラにおける射影方程式を用いてエピポーラ線の平行化を行なうための第一段階である，画像平面を投影しなおす平面の座標系を求める．

画像平面を投影し直す平面の座標系

図 5 に示す通り，新しく投影される平面 I の座標系の各 x, y, z 軸の方向ベクトルをそれぞれ $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ とすると， \mathbf{e}_1 はカメラの光学中心を結んだ基線方向と同じベクトルであり， \mathbf{e}_3 は画像平面 I_1, I_2 の交線の方向ベクトル \mathbf{L} で表される．つまり交線ベクトル \mathbf{L} は平面 I_1 の z 軸の方向ベクトル $\mathbf{r}'_{z1}{}^T$ と平面 I_2 の z 軸の方向ベクトル $\mathbf{r}'_{z2}{}^T$ の外積で求められる．そして最後に \mathbf{e}_2 はこれまでに求めた $\mathbf{e}_1, \mathbf{e}_3$ の外積で求められる．以上の事柄を式で表記すると，

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{t}_2 - \mathbf{t}_1 \\ \mathbf{e}_3 &= \mathbf{r}'_{z1}{}^T \otimes \mathbf{r}'_{z2}{}^T \\ \mathbf{e}_2 &= \mathbf{e}_1 \otimes \mathbf{e}_3 \end{aligned} \quad (8)$$

となり，各ベクトルに正規化処理を施したベクトルをそれぞれ $\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3$ とすると，投影される平面 I の座標系と世界座標系に対する各軸の相対的な関係を表す回転行列 R_I は，

$$\mathbf{e}'_1 = \frac{\mathbf{e}_1}{\|\mathbf{e}_1\|}, \quad \mathbf{e}'_2 = \frac{\mathbf{e}_2}{\|\mathbf{e}_2\|}, \quad \mathbf{e}'_3 = \frac{\mathbf{e}_3}{\|\mathbf{e}_3\|} \quad (9)$$

$$\mathbf{R}_I = \begin{bmatrix} \mathbf{e}_1^{/T} \\ \mathbf{e}_2^{/T} \\ \mathbf{e}_3^{/T} \end{bmatrix} \quad (10)$$

と表せる.

画像平面を平面 I に投影し直す変換行列

式(10)により求められた回転行列 \mathbf{R}_I により, 平面 I_1, I_2 を平面 I に投影する新しい射影行列 $\mathbf{P}'_1, \mathbf{P}'_2 (3 \times 4)$ はそれぞれ,

$$\mathbf{P}'_i = \mathbf{A}_i [\mathbf{R}_I \mid -\mathbf{R}_I \mathbf{t}_i] = [\mathbf{R}''_i \mid \mathbf{t}''_i] \quad (i = 1, 2) \quad (11)$$

となる. ここで, 世界座標系のある点 d を同次座標系で $d = (X, Y, Z, w)^T$ とすると, $w = 0$ のとき, この点 d は無限遠点を表す. この無限遠点 d と元の射影行列 \mathbf{P}_i , 式(11)で導出された射影行列 \mathbf{P}'_i から,

$$\mathbf{m}_i = \mathbf{P}_i d = \mathbf{R}'_i d \quad (i = 1, 2) \quad (12)$$

$$\mathbf{m}'_i = \mathbf{P}'_i d = \mathbf{R}''_i d \quad (i = 1, 2) \quad (13)$$

式(12), 式(13)から d を消去して, $\mathbf{m}_i, \mathbf{m}'_i$ の関係式を導くと,

$$\begin{aligned} \mathbf{m}'_i &= \mathbf{R}''_i \mathbf{R}'_i{}^{-1} \mathbf{m}_i \\ &= \mathbf{S}_i \mathbf{m}_i \quad (i = 1, 2) \end{aligned} \quad (14)$$

となる. 但し, $\mathbf{S}_i = \mathbf{R}''_i \mathbf{R}'_i{}^{-1}$ とする. この行列 \mathbf{S}_i が各ステレオ画像平面 I_1, I_2 を, 平面 I に投影し直す変換行列である.

以上の処理により, 撮影されたステレオ画像のエピポーラ線を平行化するための変換行列が得られる. 平行化を行なった各ステレオ画像をもとに, ステレオ画像の対応付けを行ない奥行き情報を求める.

3.1.3 ステレオ画像の対応付け

ステレオ視により奥行き情報を算出するためには、右の画像の中に見える点が左の画像の何処に見えるか探し出す、対応点探索問題が重要となる。この対応点を探索する方法としては大きく分けて、

- 領域に基づく方法 (area-based method)
- 特徴に基づく方法 (feature-based method)

の2つに大別される [24]。領域に基づく方法は、一方の画像のある点の対応を他方の画像から探す際、その点の周りの局所的な濃度パターンを手がかりに探索しようというものである。一方、特徴に基づく方法は、画像からエッジ等の特徴を抽出し、両ステレオ画像間の特徴間で対応付けを行なおうとするものである。

一般に、領域に基づく方法は、周辺画素の情報など、より多くの情報を用いて対応をとるため計算コストは高く、画像間の幾何学的、光学的歪みに弱いという欠点がある反面、密な視差マップが得られ、ハードウェアによる高速化に向いている。一方、特徴に基づく方法では、抽出した特徴領域に対してのみ視差が求まるため、全体的に疎な視差マップとなるが、画像間の幾何学的、光学的歪みに対して、その影響を受けにくいという利点もある。

本研究では、領域に基づく方法と特徴に基づく方法、双方の利点を生かした手法で視差マップを求める手法を提案する。具体的には、画像間の特徴領域として、画像の輝度変化が大きい部分、つまりエッジ領域を採用する。この理由として、もし画像上で輝度変化が全くなければ、当然対応を取ることは不可能であり、視差を求めることはできない。あるいは、完全に一定でなくても、輝度変化が非常に小さい領域では、その対応は求められたとしても、不確かなものでしかないからである。そこで特徴領域抽出の段階で、輝度変化が大きな所のみ抽出しておけば、輝度変化が緩やかな部分等は対応点探索時に視差推定の対象から除かれていることになる。そして、実際の対応点探索は領域に基づく方法で行なう。つまり、輝度変化が大きい領域内はブロックマッチングと呼ばれる手法を用いて、密に視差を求める。また、より信頼性の高い探索を行なうため、画像の輝度情報だけでな

く、色情報を用いた対応点探索を行う。以下に対応点探索における処理手順について述べる。

特徴領域抽出方法

入力画像の特徴領域を抽出するために、画像を輝度変化の大きな部分と小さな部分に分類する。この処理手順は以下の通りである。

1. Sobel フィルタによるエッジ部分の抽出
2. 1. の出力結果である濃淡画像を 2 値画像に変換

この処理により得られる結果を特徴領域とする。1. の処理により輝度変化の大小を濃度値で表した濃淡画像が出力される。この処理結果に対して、2. の処理により、ある一定閾値を境に 2 つの領域に分割し、輝度変化の急変する部分を特徴領域として抽出する。

このとき使用する閾値により、処理結果は大きく変化してしまう。このような閾値は本来、与えられる画像データに応じて決定されるべきものである。本研究では、この閾値を決定するために画像の濃度ヒストグラムを用い、このヒストグラムをある値で 2 つに分割し、その値を閾値として設定する。閾値を設定する際の基準としては以下の条件が挙げられる、

- 分割されたヒストグラムがそれぞれまとまりが良いこと。
- 分割された 2 つのヒストグラム間で濃度の差が大きいこと。

この条件を満たす閾値決定処理として、ここでは判別分析法 [25] を用いる。この処理で求めた閾値により濃淡画像から 2 値画像に変換し、視差の探索領域とする輝度変化の大きな部分と、輝度変化の小さな部分に分割する。

対応点探索手法

ステレオ視において、基準画像の中に見える点が参照画像のどこに見えるか探し出す対応点探索問題の手法として、画像の輝度値を特徴量としたテンプレートマッチングを行なうことにより、対応点を発見し、視

差を求める。この対応点探索を先に述べた，特徴領域内の全点について行えば，その画像対における視差マップが得られ，この視差マップを式(1)に適用することにより，そのシーンにおける奥行き情報を獲得することができる。

本研究で用いたテンプレートマッチングの評価基準には両ステレオ画像の色合いの違いを考慮して，参照テンプレート，探索テンプレート内の各輝度値を表したベクトルの正規化相互相関を用いた。この評価基準は計算コストは高くなるが，画像の線形な輝度変化に対して安定した結果を得ることができるという特徴がある。以下に，この正規化相互相関について詳しく説明する。

正規化相互相関

基準画像 I_r 上の注目点に関して，照合を行なう領域(画素位置の集合)をウィンドウ W ，テンプレートとなる部分画像の画素単位の輝度値を $f(x, y) ((x, y) \in W)$ とする。また，対応点探索を行なう参照画像 I_s 上の，ある位置 (u, v) のウィンドウによって得られる部分画像の画素単位の輝度値を $g(x, y; u, v)$ で表し，対応点の探索範囲 S はウィンドウの位置 (u, v) の集合とする。 (u, v) は画素単位の座標として表され，その場合 $g(x, y; u, v) = I_s(u + x, v + y)$ となる。ここで，マッチング処理を行なうウィンドウ W の大きさを $M \times N$ の矩形領域とすれば，正規化相互相関 $Corr(f, g; u, v)$ は以下のように表せる。

$$Corr(f, g; u, v) = \frac{\sum_{y=1}^N \sum_{x=1}^M (f(x, y) - \bar{f}) \cdot (g(x, y; u, v) - \bar{g}(u, v))}{\sqrt{\sum_{y=1}^N \sum_{x=1}^M (f(x, y) - \bar{f})^2} \sqrt{\sum_{y=1}^N \sum_{x=1}^M (g(x, y; u, v) - \bar{g}(u, v))^2}} \quad (15)$$

ここで， $\bar{f}, \bar{g}(u, v)$ はウィンドウ内の画素値の平均値であり，

$$\bar{f} = \frac{1}{NM} \sum_{y=1}^N \sum_{x=1}^M f(x, y) \quad (16)$$

で与えられる。 $\bar{g}(u, v)$ も同様であるが、この値はウィンドウの位置 (u, v) に依存する。この正規化相互相関値は、 -1 から 1 の範囲をとり、値が大きいほど画像が類似していることを表す。したがって、正規化相互相関を用いた対応点探索は $Corr(f, g; u, v)$ を最大とする (u, v) を探索範囲から見つける処理となる。

正規化相互相関は先に抽出された特徴領域全体に対して行なわれる処理であり、たとえ計算された正規化相互相関値が低くても、その時の対応点が視差として求められることになる。このことは求められた相関値が低い場合、それだけ信頼性の低い対応点候補ということになり、実際に対応する点とは異なった対応付けをしている可能性が高い。また、もともと撮影場所の違う画像のため、一方の画像には写っていても、他方の画像には写っていないというオクルージョンの問題や照明状況などの影響により対応点自体存在しない場合も考えられる。そこで、ある一定閾値以下の相関値の場合は、その値は視差マップに含めないようにする。この処理を施すことにより、対応点の相関値が大きい、信頼性の高い視差マップを作成する。

色情報への拡張手法

テンプレートマッチングにおける評価基準について、入力画像の輝度情報だけでなく、色情報も重要な情報源と考え、ここでは入力画像を HSV 色空間へ投影し、輝度 (V) と色相 (H) の双方を用いたマッチングを行う手法を提案する。HSV 色空間の特徴として色相は彩度 (S) が小さい場合には情報量は少なく、逆に彩度が大きい場合、色相は多くの情報量を持つので、実際には以下のような評価関数 d をマッチング尺度として用いる。

$$d = \{S \cdot H + (1 - S) \cdot V\} \quad (17)$$

$$\text{但し, } \begin{cases} H : \text{色相情報での正規化相互相関値} \\ V : \text{輝度情報での正規化相互相関値} \\ S : \text{彩度情報 (0 \sim 1)} \end{cases}$$

この評価関数 d も -1 から 1 の値を取り、値が大きいほど画像が類似していることを示している。この評価基準について、各色情報は HSV 六角錐

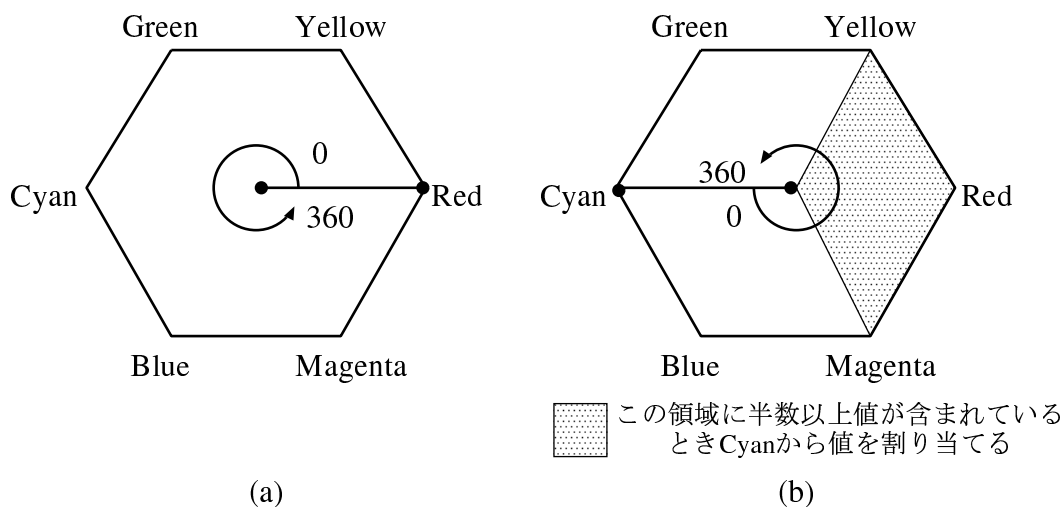


図 7 HSV6 角錐カラーモデル

カラーモデル [26] に基づいている．色相値は連続しているが，実際には図 7(a) のように数値で表される．そのため，Red 領域付近において，値は連続しているにも関わらず，色相値が全く掛けはなれた値になるという問題が生じる．この問題を回避するため，照合を行なう領域のウィンドウに含まれる値が，色相情報の Red を中心に ± 60 度以内の値を半数以上持つ場合には，図 7(b) のように Red の補色である Cyan から数値を割り当て，Red 領域の連続性を保つ処理を行う．

以上の処理が，ステレオ視による奥行き情報の獲得手法である．この処理により，一般的なステレオ撮像系で撮影された画像に対しても，効率良く対応点を探索でき，また，特徴領域の奥行き情報のみ求めているので，計算コストの低い，かつ信頼性の高い奥行き情報を得ることができる．

この処理により得られる奥行き情報は，エピポーラ線を平行化するために変換を行なったステレオ画像に対して得られるものである．最終的な奥行き情報は，撮影された入力画像の位置・姿勢の関係に合わせるため，基準入力画像に対しての奥行き情報を必要とする．そこで，エピポーラ線の平行化処理を行なった変換行列の逆行列を求め，その逆行列を，得られた奥行き画像に適応することで，撮

影時の入力画像に対する奥行き情報を得る。

3.2 奥行き情報からの平面の推定

本手法のように、特徴領域の視差推定から奥行き情報を獲得しても、その奥行き情報はあくまで、特徴領域のみの疎な視差マップにすぎない。視差から求まる奥行き情報も同様に、疎な奥行き情報にすぎず、特徴領域のみの奥行き情報から、奥行き情報が求まっていない部分を求めなければならない。

従来、部分的な奥行き情報から、奥行き情報が求まっていない部分を求める方法としては、得られている奥行き情報からの補間で行なわれているか [8, 27], もしくはあらかじめ同一の平面を構成する対象領域の情報を手動で付加し、その与えられた領域の中に含まれる奥行き情報から、対象領域の奥行き情報を抽出するという方法が取られていた [10]. しかし、奥行き情報を単純に補間していたのでは、誤対応などから起こる奥行き情報の影響を、直接受けてしまう可能性がある。特に屋外環境下でのステレオ撮影は、カメラ位置姿勢の厳密な測定は難しく、複雑な照明環境下では画像に多くのノイズなどが含まれ、そのような状況下において、特徴領域の中に含まれている点全てが、正確に対応付けされているとは限らない。また、あらかじめ手動で領域情報を付加する手法は、正確に対象領域の3次元平面を推定することが可能だが、再構成する環境が大規模になればなるほど、ユーザにかかる負担は大きくなる。

本研究は自動で、かつ密な奥行き情報を獲得するために、入力画像上で同一の平面を構成する対象領域の情報を自動で抽出し、その抽出された領域の中に含まれる奥行き情報から、対象領域の奥行き情報を抽出する手法を提案する。本提案手法は図 8 に示すように、前処理として入力画像に同一平面を構成する領域毎に分割処理を施す (図 8(a)). この処理により抽出された領域をそれぞれ一つの3次元平面を構成すると仮定し、その抽出された領域に存在する奥行き情報から、各領域毎の法線ベクトルを求め (図 8(b)), 実際の3次元平面を推定する。推定された3次元平面を3次元空間に投影し、撮影された現実環境における幾何形状を描画する。以下、各処理について詳細に説明する。

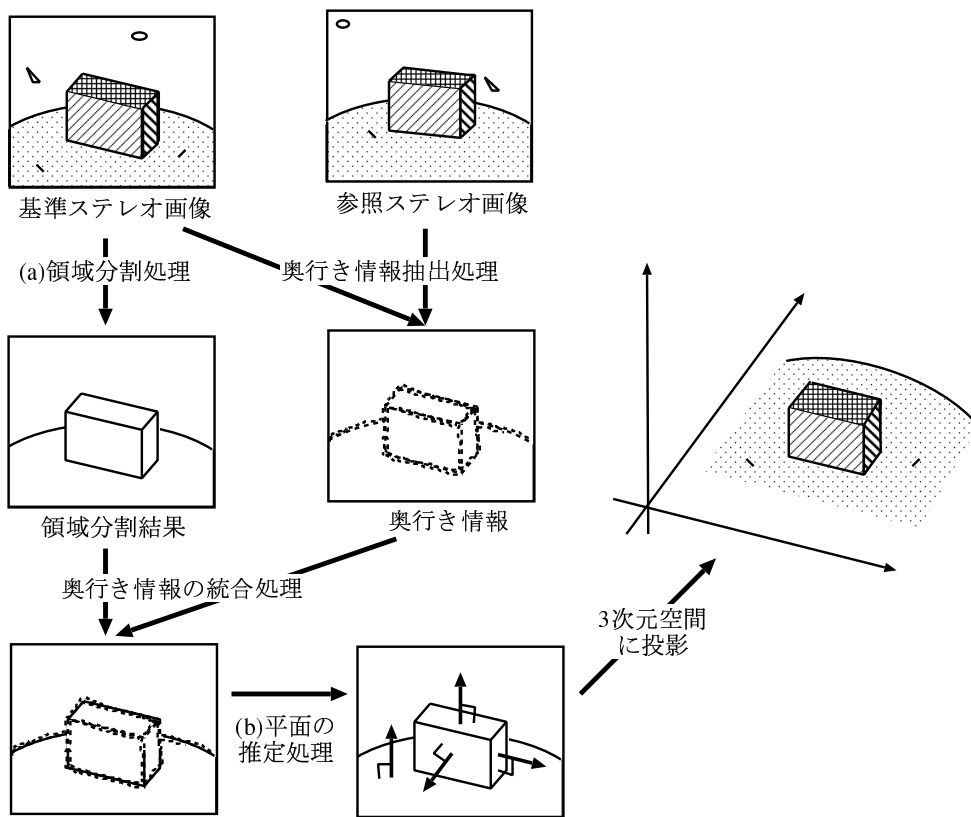


図 8 平面推定における処理の流れ図

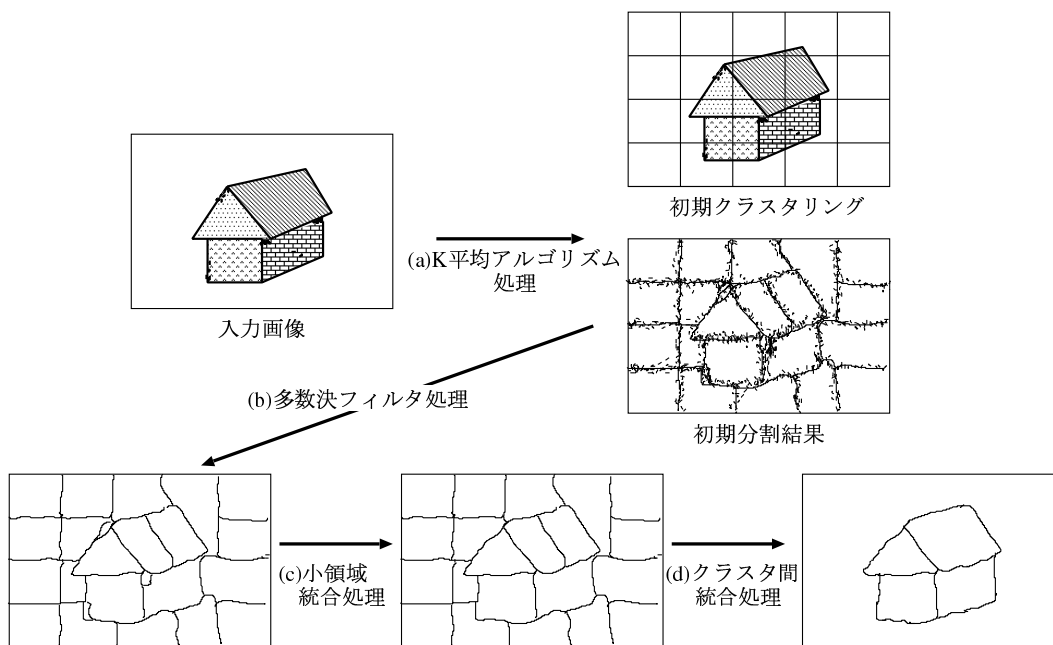


図 9 領域分割処理手順

3.2.1 画像の領域分割

入力画像上で同一の平面を構成する対象領域の情報を自動で抽出するため、本節では入力画像の領域分割手法について説明する。画像の領域分割の方法には大きく分けて、画像の輝度値やエッジ部分などの特徴空間を用いた方法、または、画素の隣接関係などの画像空間を用いた方法の2種類がある。特徴空間を用いた方法はノイズに強く、大まかな領域を分割するのに適しているが、細かい領域を分割するには適していない。一方、画像空間を用いた方法は詳細な領域を分割するのに適しているが、ノイズに弱く、大まかな領域をひとまとめにして分割するには不向きである。このように各手法は相補的な関係にある。本研究では、各領域の境界は厳密に分割され、かつ、一つの平面が複数の領域に分割されるのではなく、一つの領域として分割することを目指す。そこで、2種類の手法の各相補的關係に着目し、特徴空間と画像空間の併用による領域分割法を提案する。

最初に、大まかな領域分割を行なうため、画素の輝度値などの特徴量を用いたK平均アルゴリズム処理 [28] を行なう (図 9(a))。しかし、このK平均アルゴリズムにより得られる領域分割結果には過分割による小領域が大量に発生してしまうという問題がある。提案手法ではこの小領域の問題を回避するため、多数決フィルタ処理を行なうこと (図 9(b))、また、ある一定画素数以下の領域は、その領域に隣接する画素数の最も多い領域に併合する処理を行なうことにより (図 9(c))、小領域を統合する。このようにして得られた大まかな領域分割結果に、各クラス間同士の統合処理を行ない (図 9(d))、最終的な領域分割結果とする。以下に各処理手順を示し、各処理について説明する。

1. K平均アルゴリズムによるクラスタリング処理
2. 小領域の除去処理
3. クラスタ間境界情報と色情報に基づくクラスタ間の統合処理

K平均アルゴリズムによるクラスタリング

K平均アルゴリズムとは初期状態として適当なクラスタを与え、そのクラスタを組み変えて少しずつ、「より良い状態」を求めていくものである。こ

のアルゴリズムは複数の特徴量を一元的に取り扱うことができるという利点がある。この手法は以下の手順で実現される。

1. 初期状態として、画像平面上を矩形に分割したクラスタを与える。分割するクラスタ数は予め設定しておく。例えば入力画像を横 M 分割、縦 N 分割に区切ったとすると、初期クラスタは $M \times N$ 個のクラスタとなる。2. の処理へ。
2. 各クラスタ毎に特徴量の平均値を計算する。ここで用いる特徴量については後で詳しく述べる。3. の処理へ。
3. 全画面の画素ごとにその画素の特徴量がどのクラスタの特徴量の平均値と近いかを、後述する式 (18) の距離によって調べ、そのクラスタに統合することでクラスタを再構成する。4. の処理へ。
4. 再構成前後で、クラスタ内の平均値が変化するとき、2. の処理へ戻る。クラスタ内の平均値が変化しなければ、クラスタリング処理を終了する。

2. の処理におけるクラスタリングの特徴量として、ここでは、画像の色情報だけでなく、画像の特徴的な性質ともいえる連続性も考慮する。その理由として、輝度情報や色情報の特徴量のみでクラスタリングを行なう場合、領域境界が不連続になったり、境界領域と実際の画像のエッジの位置が一致しない部分が生成されてしまい、領域分割の精度が劣化するという問題が生じるからである。そこで、特徴量空間に画面上の位置情報を加えた多次元特徴空間内でクラスタリングを行なうことでこの問題を回避する。しかし、位置情報を加えることにより、画像の連続性は考慮されるが、大きな領域は複数の領域に分割されてしまう。そこでもともと一つの領域であるにもかかわらず、複数に分割されたクラスタの統合処理を行なう必要がある。この処理の詳細は後で述べる。

この位置情報と色情報を用いた特徴空間として RGB 色空間の 3 次元と HSV 色空間上での評価関数値の 1 次元の合計、4 次元の特徴量と入力画像の水平

垂直位置 (x, y) の 2 次元の特徴量との合計 6 次元特徴空間を用いる。その時のクラスタ間の距離を

$$\sqrt{k_0\{(R - \bar{R})^2 + (G - \bar{G})^2 + (B - \bar{B})^2 + (d - \bar{d})^2\} + k_1\{(x - \bar{x})^2 + (y - \bar{y})^2\}} \quad (18)$$

とし、3. の処理における判定基準とする。ここで、 R, G, B, x, y は画素の特徴量、 $\bar{R}, \bar{G}, \bar{B}, \bar{x}, \bar{y}$ はクラスタの平均値であり、 k_0, k_1 は特徴量への重み付け係数である。各特徴量は全て正規化されており、 d に関しては、前述した式 (17) の評価関数値を用いる。

小領域の除去

K 平均アルゴリズムで大まかな領域分割処理を行なうが、このアルゴリズムの領域分割結果には過分割によるごましお状の意味のない領域が大量に生成されてしまうという問題がある。そのため、この微小領域を取り除くために、2 つの処理を行なう。1 つ目は、多数決フィルタ処理を用いること、そして 2 つ目は一定画素数以下の領域は、その領域に隣接する画素数の最も多い領域に併合する処理を行なうことである。

多数決フィルタは図 10 のように中心画素を除くウィンドウ内で最も画素数の多いカテゴリーで中心の画素のカテゴリーを置き換えるというものであり、単純な処理によって粒状の微小領域を大きな領域に統合することができる。

多数決フィルタにより、微小領域の統合が行なわれるが、このフィルタにより統合されなかった小領域に関しても、ノイズの影響や、シーン中において主となる平面以外の領域の可能性が高い。そこで、ある一定画素数以下の領域は隣接領域に統合する処理を行なう。統合する隣接領域は、統合される小領域の最外郭の画素を統合領域判定の対象画素とし、その対象画素が隣接する別領域に最も多く属している領域を統合領域とする。

クラスタ間境界情報と色情報に基づくクラスタの統合

これまでの処理で、クラスタ間の境界において細かいクラスタリングが行

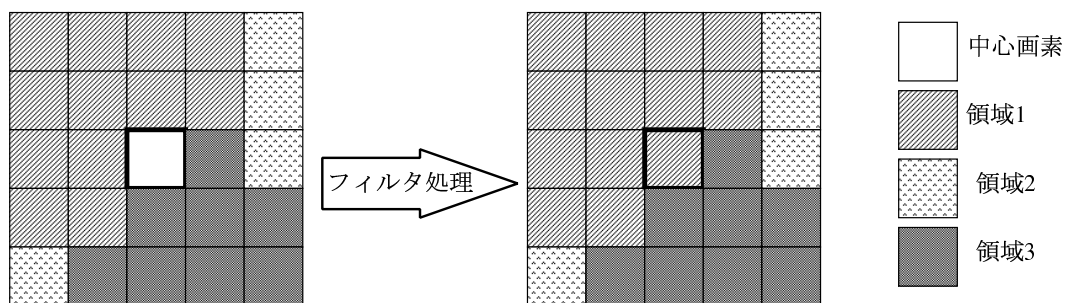


図 10 多数決フィルタの例

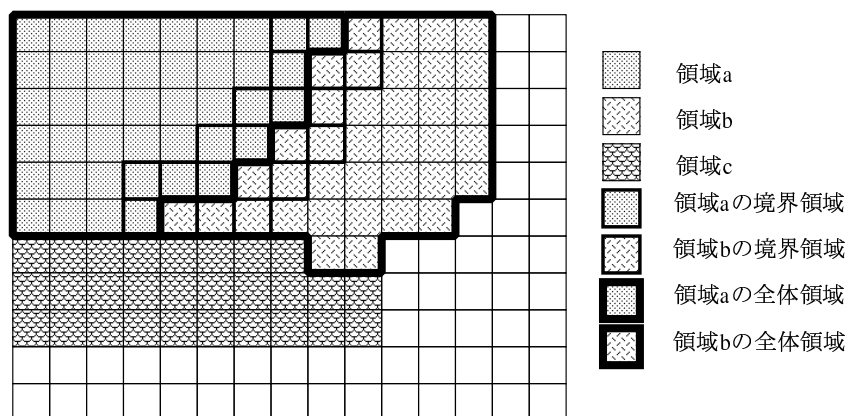


図 11 クラスタ間統合説明図

なわれる反面、大きな領域は複数の領域に分割されてしまうという問題が生じる。そこでもともと一つの領域であるにもかかわらず、複数の分割されたクラスタ間の統合処理を行なう必要がある。以下では、この統合処理について説明する。

もともと、一つの独立した領域の境界は、色情報の変化が大きい傾向にある。一方、一つの大きな領域が複数の領域に分割された場合の境界上の色変化は小さい。この性質を考慮した統合手法として、画面上で隣接するクラスタ間の境界での色情報の変化が小さい所を統合する処理を行なう。具体的なクラスタ間の距離の評価基準は、

$$\sqrt{k_0\{(\bar{R}_a - \bar{R}_b)^2 + (\bar{G}_a - \bar{G}_b)^2 + (\bar{B}_a - \bar{B}_b)^2 + (\bar{d}_a - \bar{d}_b)^2\} + k_1\{(\bar{R}_{ra} - \bar{R}_{rb})^2 + (\bar{G}_{ra} - \bar{G}_{rb})^2 + (\bar{B}_{ra} - \bar{B}_{rb})^2 + (\bar{d}_{ra} - \bar{d}_{rb})^2\}} \quad (19)$$

とする。図 11に示すように、領域 a の全体領域の RGB の平均値をそれぞれ $\bar{R}_a, \bar{G}_a, \bar{B}_a$ 、領域 b の全体領域の RGB の平均値をそれぞれ $\bar{R}_b, \bar{G}_b, \bar{B}_b$ 、また、領域 a の境界領域の RGB の平均値をそれぞれ $\bar{R}_{ra}, \bar{G}_{ra}, \bar{B}_{ra}$ 、領域 b の境界領域の RGB の平均値をそれぞれ $\bar{R}_{rb}, \bar{G}_{rb}, \bar{B}_{rb}$ とする。また、 d は式 (17) で表される評価関数値であり、 k_0, k_1 は各特徴量への重み付け係数である。この評価基準の各値は正規化されており、この距離が一定閾値以下なら統合処理を行なう。この処理を隣接する全てのクラスタ間同士で行ない、統合できる領域がなくなったとき終了する。

3.2.2 平面の推定

3.2.1章で述べた領域分割結果から得られた各領域がそれぞれ一つの 3 次元平面を構成すると仮定し、その抽出された領域に存在する奥行き情報から、各領域毎の法線ベクトルを求め実際の 3 次元平面を推定する。平面の推定方法として、ここでは統計的手法を用いた平面方程式の推定を行なう。

統計的手法は、実際のデータ（ここでは奥行き情報から求められた 3 次元データ）と、そのデータの間を抽象化したモデルとの間の確率統計的な関係を解析する手法である。この代表的な手法として最小 2 乗法やロバスト推定 [29, 30] が挙げられる。

最小 2 乗法は、データにモデルを当てはめた際の誤差の 2 乗和を最小とするようにモデルのパラメータを決定する手法であり、コンピュータビジョンの分野でも広く利用されている。しかし、モデルの推定方法として最小 2 乗法を利用する場合、幾つかの問題が生じてくる。その問題とはモデルとデータとの誤差が平均 0 の正規分布に従う場合には、推定されたモデルは最適となるが、データに例外値が含まれているような場合には、推定結果が例外値の影響を大きく受けてしまう。図 12(a) に示すように、最小 2 乗法により直線のモデルをデータに当てはめ

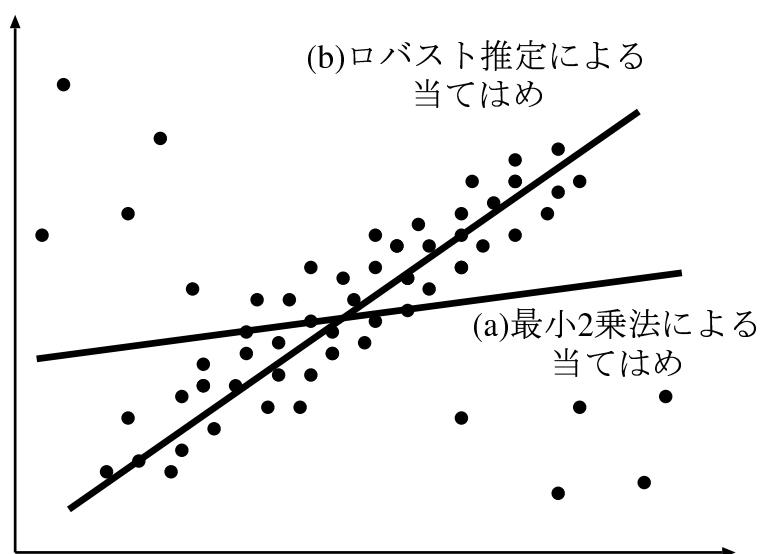


図 12 ロバスト推定

た場合、例外値の影響を直接受けた結果となる。

そのような場合には、データ内の主要な部分のみモデルを当てはめ、残りの部分を例外値として自動的に除外すれば、求めるパラメータを精度良く推定できる。この統計的手法として、ロバスト推定がある。図 12(b) に示すように、ロバスト推定により直線のモデルをデータに当てはめた場合、例外値の影響を抑えた推定結果となる。

本研究では、ステレオ視により得られる 3 次元データが、必ずしも全て正しいデータとは限らない。特に屋外環境下において撮影されたステレオ画像から、視差を求め、3 次元データを獲得する処理の過程において、計測された全てのデータが正確な値である可能性は低く、不正確な値も含まれている。最小 2 乗法で平面方程式のパラメータを推定する場合、このような誤った値も含めて平面方程式のパラメータを推定してしまうという問題が生じてしまう。そこで、ステレオ視により得られる 3 次元データのうち、誤対応による不正確な 3 次元データを外れ値とみなし除去する手法としてロバスト推定を用いた平面方程式の推定を行なう。

ロバスト推定では得られたデータに外れ値が半分以上含まれている場合、正し

い値の推定はできないが，ステレオ視により得られる3次元データのうち，誤対応等の原因により外れ値となる3次元情報は半分以下となる仮定のもとでロバスト推定を適用する．ロバスト推定を用いると，例外値をある程度含むようなデータからでも比較的安定にモデルのパラメータを推定できる．このロバスト推定のための代表的な手法としては，LMeds 推定，LTS 推定，M-推定²がある．

3つの推定方法のうち，M-推定法は高い推定精度が得られるが，このM-推定は非線形最適化を用いて行なわれるため，初期値の設定，適切な収束範囲の設定など，その実現に際して困難な部分も多い．次に LMedS 推定と LTS 推定であるが，いずれも残差のソーティングが必要である．しかし，LTS 推定の場合さらに2乗処理が必要であるため計算量が若干増加する上，LMedS 推定と比較して特に推定誤差が改善されるわけではない．さらに，LMedS 推定はランダムサンプリングによって計算可能であり，比較的安定な処理が期待できる．このような理由から LMedS 推定を用いた平面方程式の推定を行なう．以下，LMedS 推定を用いた平面方程式推定のアルゴリズムを図13に沿って説明する．

1. 対象領域内に存在する3次元情報からランダムに3点を選択する (図13(a))
2. 選択した3点から，1つの平面方程式を生成する．例えば，図13(b)のように対象領域内の点 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ が選択され， $\mathbf{m}_1 = \mathbf{a}_2 - \mathbf{a}_1$ ， $\mathbf{m}_2 = \mathbf{a}_3 - \mathbf{a}_1$ とすると，求める平面の法線ベクトル \mathbf{n} は $\mathbf{n} = \mathbf{m}_1 \times \mathbf{m}_2$ となる．この法線ベクトル \mathbf{n} と点 \mathbf{a}_1 を通る平面上の点を \mathbf{p} とすると，点 \mathbf{p} は， $\mathbf{n} \cdot (\mathbf{p} - \mathbf{a}_1) = 0$ を満たす平面上に存在する．この平面が3点 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ を通る平面方程式となる．
3. 算出した平面方程式を用いて，領域内に存在する3次元情報全てを算出した平面に投影し，投影後の点と実際に算出された3次元情報との差分ベクトルを計算する (図13(c))
4. 差分ベクトルの大きさのメディアン値を計算する (図13(d))

²各推定方法の詳細な説明は付録Bで述べる

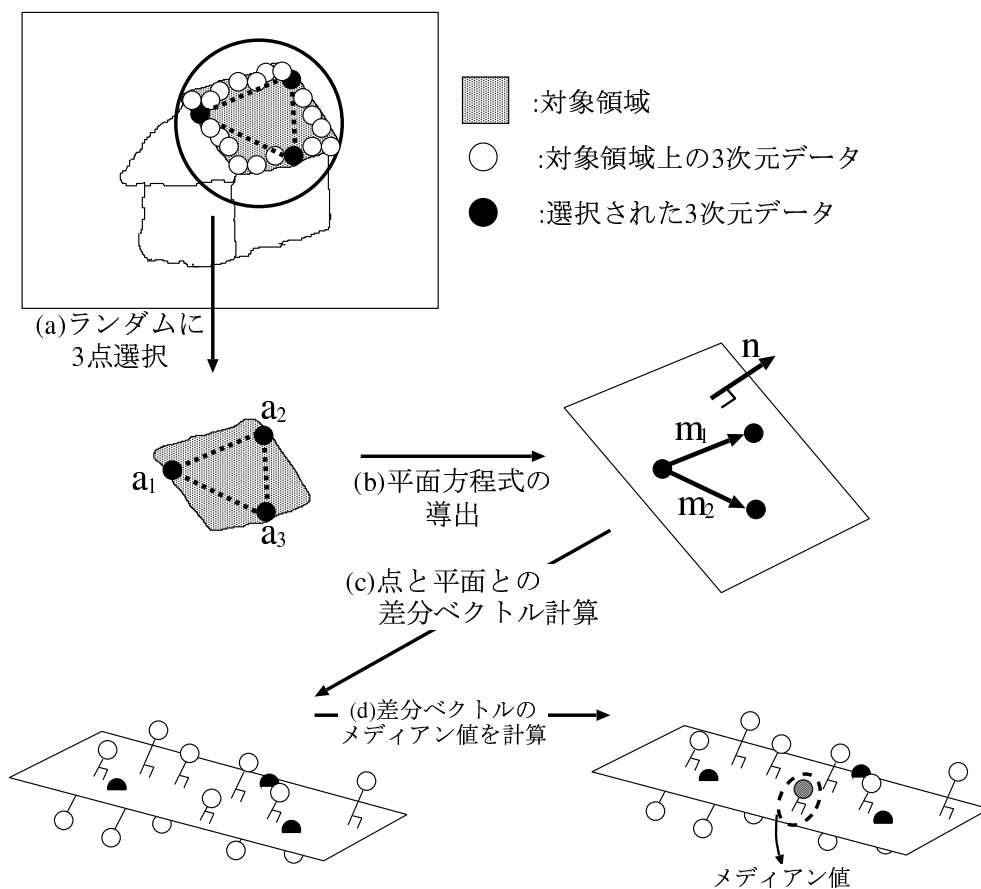


図 13 平面推定アルゴリズムの説明図

5. 1. ~ 4. を k 回繰り返して差分ベクトルの大きさのメディアン値が最小となる時に導出した平面方程式を、その領域の平面方程式とする

ここで、平面の推定にロバスト推定を用いる場合、推定する平面上に存在する奥行き情報が極端に少ない場合、ロバスト推定を行っても信頼性の高いパラメータを得ることはできない。例えば、ある平面を推定する場合、その平面上に存在する奥行き情報が3点しかない場合、求められる平面はその3点で構成される平面になってしまい、このままでは正しく推定されているかどうか判断できない。このような理由から、平面を推定する場合、その平面上にのっている奥行き情報が、ある一定個数以内の場合は、平面の推定を行わないようにする。平面

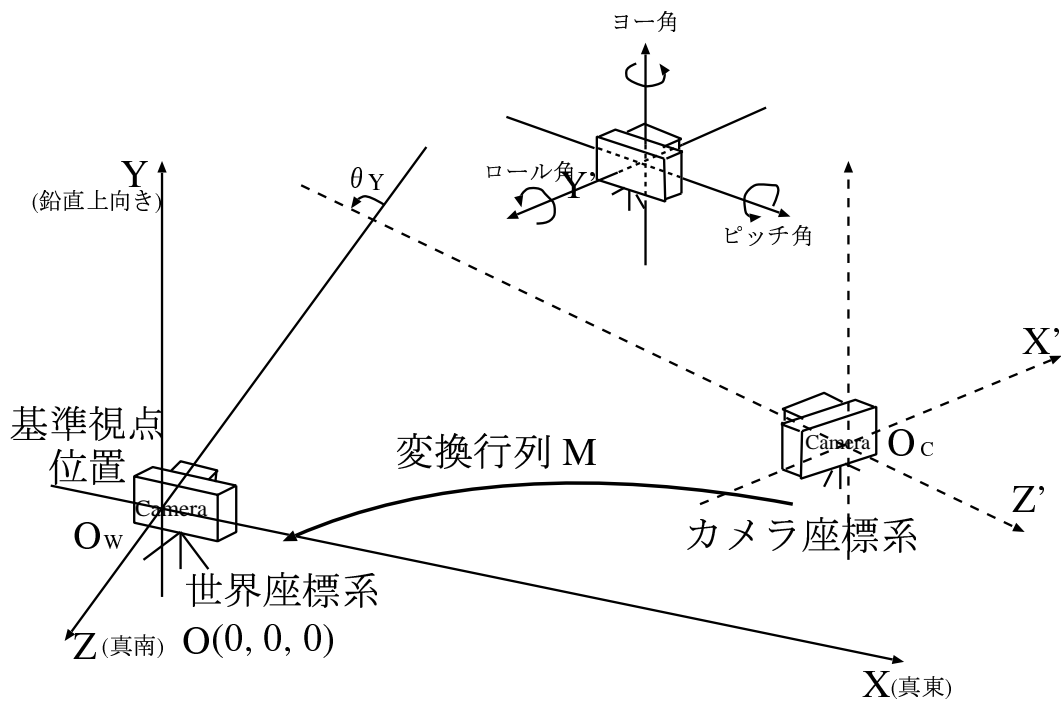


図 14 カメラ座標系と世界座標系間の関係

推定が行なわれなかった領域については、3次元モデルを再構成する場合にその領域は描画しない。

また、ロバスト推定の試行回数については4章で詳しく説明する。

3.3 多視点データの統合

これまでの章において、各撮影位置からの3次元モデルを生成する処理手順について述べてきた。本章では各撮影位置での3次元モデル、つまり各々のカメラ座標系で構成される3次元モデルを、ある一つの基準座標系(世界座標系)へと変換し、各々の3次元モデルを統合する手法について述べる。

各カメラ座標系から世界座標系への変換は、各撮影地点でのGPSと3軸角度センサにより得られる計測データを基に行う。図14のように、基準となる撮影位置におけるカメラ座標系の原点を世界座標系の原点 O_w に一致させ、世界座標系

の X, Z 軸は絶対水平面に一致させる．また X 軸正の方向を真東， Z 軸正の方向を真南とし， Y 軸は鉛直上向きを正とする．この時，各視点から得られた奥行き情報は各視点でのカメラ座標系で表現されており，各カメラ座標系から世界座標系への変換行列が必要になる．以下，3次元座標を同次座標系で表現すると，ある点の世界座標系の座標値 \mathbf{x} は，カメラ座標系での座標値 \mathbf{x}' と変換行列 \mathbf{M} を用いて，

$$\mathbf{x} = \mathbf{M}\mathbf{x}' \quad (20)$$

と表現できる．この時，カメラ座標系の世界座標系に対する相対的な位置を $\mathbf{t} = (t_x, t_y, t_z)^T$ とし，絶対水平面を基準としたカメラ座標系の傾斜角 (ピッチ角，ロール角) を θ_P, θ_R ，及び子午線を基準とした方位角 (ヨー角) を θ_Y とすると，カメラ座標系から世界座標系への回転行列 \mathbf{R} は，各軸の回転行列の積として式 (21) のように表される．

$$\mathbf{R} = \mathbf{R}_P(X' : -\theta_P)\mathbf{R}_Y(Y' : -\theta_Y)\mathbf{R}_R(Z' : -\theta_R) \quad (21)$$

$$\mathbf{R}_P(X' : -\theta_P) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(-\theta_P) & -\sin(-\theta_P) & 0 \\ 0 & \sin(-\theta_P) & \cos(-\theta_P) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_Y(Y' : -\theta_Y) = \begin{bmatrix} \cos(-\theta_Y) & 0 & \sin(-\theta_Y) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(-\theta_Y) & 0 & \cos(-\theta_Y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}_R(Z' : -\theta_R) = \begin{bmatrix} \cos(-\theta_R) & -\sin(-\theta_R) & 0 & 0 \\ \sin(-\theta_R) & \cos(-\theta_R) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

ここで， $\mathbf{R}_P(X' : -\theta_P), \mathbf{R}_Y(Y' : -\theta_Y), \mathbf{R}_R(Z' : -\theta_R)$ はそれぞれカメラ座標系の各軸を世界座標系の各軸に合わせるための回転行列である．また，平行移動行列

を \mathbf{T} とすると, \mathbf{T} は式 (22) と表される.

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (22)$$

変換行列 \mathbf{M} は, 式 (22) の平行移動行列 \mathbf{T} と式 (21) により導出された回転行列 \mathbf{R} を用いて,

$$\mathbf{M} = \mathbf{RT} \quad (23)$$

と表される. この変換行列を用いることにより, 各視点から得られた複数のカメラ座標系を, 世界座標系に統一して表現することができる.

4. 実験

本章では、各センサにより撮影位置間の位置・姿勢の関係が既知な多視点からの入力画像を取得し、本提案手法を用いて屋外環境構造を再構成する処理過程を示すとともに、最終的な3次元モデルを示す。

4.1 実験環境

実験を遂行するに当たり各撮影位置間の幾何学的関係、つまり各カメラ間の位置、姿勢データの計測装置として位置センサと3軸角度センサを用いた。カメラ間の位置センサにはGPS(NovAtel社:RT-20(図15))を使用した。このGPSは5Hzのサンプリングレートで $\pm 100\text{mm}$ の精度の位置測定が可能である。姿勢センサには3軸角度センサ(データ・テック社:ジャイロミニ dt No.4-1(図16))を使用した。このセンサの精度はロール角, ピッチ角, ヨー角(方位角)全て ± 0.5 度であり, サンプリングレートは30Hzである。

これらの各センサを用いて、実際に屋外環境構造の撮影を行なった。各撮影カメラ間の関係と各センサにより得られた実測データをそれぞれ図17と表1に示す。ここで、GPSにより得られるデータは図17の撮影地点1を原点とする世界座標系に対する3次元位置データを出力する(単位:mm)。この世界座標系の各軸に関して、Z軸は原点Oを基準として子午線に一致する軸とし、原点Oから真南に向かう方向を正とする。X軸はZ軸に直交する軸とし、真東に向かう方向を正とする。また、Y軸をX,Z軸に垂直な軸とし、鉛直上向きの方向を正としている。次に、3軸角度センサにより得られる3次元姿勢データは絶対水平面を基準としたカメラの傾斜角、及び真北の方向を基準とした方位角データを出力する(単位:度)。カメラの傾斜角(ロール角, ピッチ角)はそれぞれ時計回り方向を正とした回転角を出力し、カメラの方位角(ヨー角)は真北方向からの回転角を出力する。

また、実験に用いた各パラメータを表2に示し、そのパラメータについて幾つか説明する。領域分割時の式(18)の重み付け係数 k_0, k_1 は、それぞれ $1/8, 1/4$ に設定している。距離に関して、このように設定することにより、濃淡値や色彩情報が急激に変化する場所においては、位置情報の変化に比べ色情報の変化の方が



図 15 GPS の外観



図 16 3 軸角度センサの外観

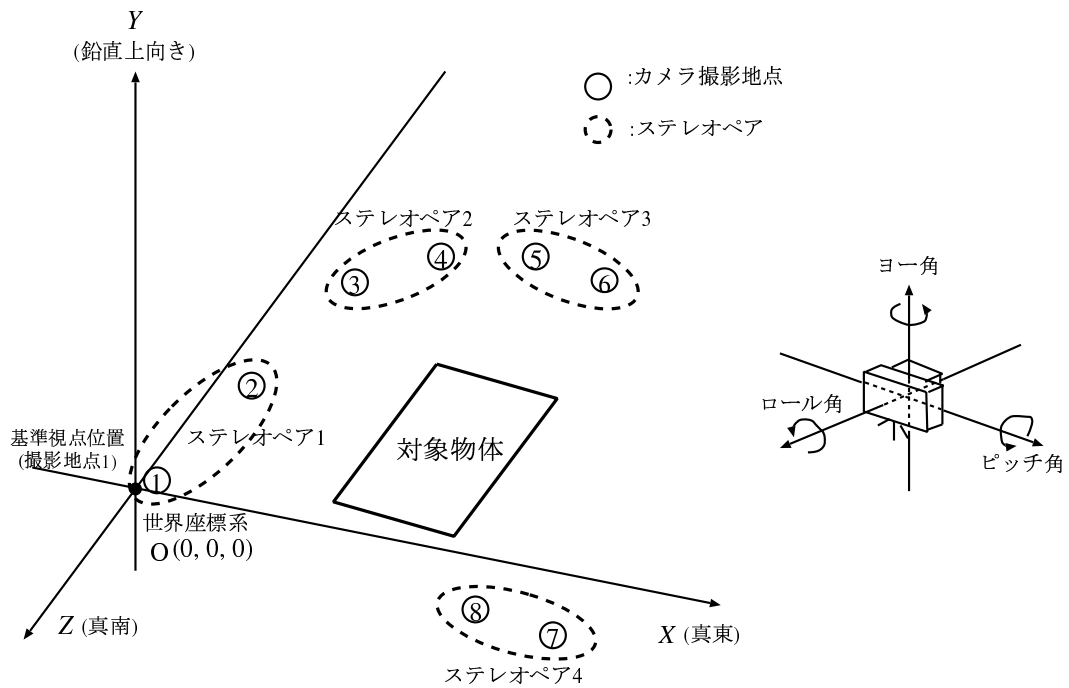


図 17 実験におけるカメラ配置

表 1 各センサの実測データ

撮影地点	X(mm)	Y(mm)	Z(mm)	ロール角 (度)	ピッチ角 (度)	ヨー角 (度)
1	0	0	0	-0.3	10.1	-83.4
2	375	-35	-4,284	0.0	10.1	-106.2
3	2,222	-58	-9,145	-1.0	11.5	-129.1
4	5,035	108	-11,919	-0.1	10.3	-148.7
5	11,482	85	-13,743	0.2	10.6	-177.2
6	16,218	132	-12,959	-0.4	12.0	162.3
7	20,007	-32	8,768	1.2	11.8	45.6
8	16,791	-65	8,299	1.5	19.0	32.2

表 2 実験において利用したパラメータ

画像サイズ	$640 \times 500 \text{ pixel}$
テンプレートサイズ	$9 \times 9 \text{ pixel}$
正規化相互相関値の閾値	0.7
K 平均アルゴリズムの初期クラス	$10 \times 10 \text{ class}$
多数決フィルタのウィンドウサイズ	5×5
式 (18) の重み付け係数	$k_0 = 1/8, k_1 = 1/4$
領域分割の小領域統合要素数	300
式 (19) の重み付け係数	$k_0 = 1/8, k_1 = 1/8$
クラスタ間の統合処理閾値	0.6
平面推定時の最低奥行き情報数	20
ロバスト推定の試行回数	100

大きくなり、画像のエッジは境界領域として保存されるという性質を持つようになる。

次に、多数決フィルタにより、微小領域の統合が行なわれるが、このフィルタにより統合されなかった小領域に関しても、ある一定画素数以下の領域は隣接領域に統合する処理を行なった。本研究では、この一定画素数の閾値を 300 画素とした。閾値である 300 画素の領域は 1 フレームの総画素数 $640 \times 500 = 320,000$ に対して約 0.1% の面積である。この程度の大きさの領域は実際の画面上で統合しても問題無いと判断した。

また、式 (19) の重み付け係数 k_0, k_1 は各々 $1/8$ に設定している。これにより、クラス間全体の色情報と境界上の色情報の重みが同等の評価基準となり、距離の値としては $0 - 1$ の値を取る。この値が一定閾値以下なら統合処理を行なう。この処理を隣接する全てのクラス間同士で行ない、統合できる領域がなくなったとき終了する。統合時の閾値は 0.6 とする。この値は経験的に求めた。

また、平面推定時のロバスト推定の試行回数は 100 とした。これは以下の理

由による。領域内において、推定される奥行き情報が誤る確率を ϵ とすると、処理 1. ~ 4. の操作を k 回繰り返した場合に、全て誤った平面が導出される確率は $(1 - (1 - \epsilon)^3)^k$ となり、正しく領域の平面方程式が導出される確率は

$$p(\epsilon, k) = 1 - (1 - (1 - \epsilon)^3)^k \quad (24)$$

と表される。式 (24) から $\epsilon = 0.4$ の場合、 $k = 100$ とすることで $p \approx 0.999$ となる。もし、 $p \approx 0.999$ であれば、平面推定の誤りは 1000 回の推定の内、1 回となり十分な精度であると思われることから本研究では k を 100 と設定した。

4.2 実験結果と考察

本手法の一連の処理結果を、図 17 のステレオペア 2 番で撮影された画像 (図 18(a)) を例にとり、その画像ペアにおけるエピポーラ線の平行化、特徴領域の抽出、視差推定処理、そして領域分割、平面方程式の推定処理結果について考察する。また、ステレオペア 1,3,4 番から得られる 3 次元モデルを統合して、対象物体の 3 次元モデル生成結果を示し、併せて考察を行なう。

【エピポーラ線の平行化】

ステレオペア 2 における撮影位置 3, 4 番目の間の幾何学的関係からエピポーラ線の平行化を行った出力結果を図 18(b) に示す。この結果、厳密なエピポーラ線の平行化までには至らず、上下数 pixel のずれが生じている。この理由として、GPS と 3 軸角度センサで計測した各カメラ間の位置、姿勢データにノイズが含まれており、このノイズのためにエピポーラ線がずれたものと考えられる。

【特徴領域の抽出】

特徴領域として抽出された結果を図 18(c) に示す。ここで、特徴領域は出力画像の白色領域である。この結果より入力画像の輝度変化の大きな部分が特徴領域として抽出されていることが確認できる。

【視差推定】

特徴領域についてのみ、視差を導出した視差マップを図 19(a) に示す。結果

の濃淡画像において、明るい点ほど近くにあることを示している。ここで、出力された視差はエピポーラ線の平行化を行なうために画像を変形させた基準画像での視差なので、変形前の画像へ逆写像を行なう(図 19(b))。この視差を基に奥行き情報を得る。

【領域分割】

図 18(a) の左画像 (基準画像) に K 平均アルゴリズムを適応した分割結果を図 20(a) に示す。この結果から分割された各領域の境界部分には、過分割による小領域が多数存在していることが確認できる。次に、この小領域を取り除くために、多数決フィルタ処理を施した結果を図 20(b) に示す。この結果により大部分の粒状の小領域が取り除かれていることが確認できる。

これまでの処理で、クラスタ間の境界において正確なクラスタリングが行なわれることがわかる。そして、その分割結果を基に、各クラスタ間を統合した結果が図 20(c) となる。この結果からもともと一つの領域であるにも関わらず、空間的拘束により複数の領域に分かれていた領域も統合されていることが確認できる。

【平面の推定】

図 19(c) に示すように、領域毎に奥行き情報が密に求まっていることが確認できる。また、図 19(b) で奥行き情報が求まっているにも関わらず、平面が求まっていない領域は、平面推定における奥行き情報の数が閾値以下であったり、奥行き情報が矛盾して平面が求まらなかった領域である。

得られた奥行き情報と領域分割結果により、領域毎に推定された平面を用いて、実際の 3 次元モデルを作成した様子を図 21(a) に示す。この図は撮影位置 3 の 3 次元モデルとなっている。また、図 21(b) に、ワイヤースケルト表示を示す。出力結果より、正確に再構成されている領域もあるが、奥行き情報のノイズや、平面推定の失敗などの原因により正確に再構成されていない領域も存在する。例えば図 21の結果には、雲領域が描画されているが、実際の位置とは明らかに異なる。この理由としては、対応点探索時に正確な対応付けができていないことや、

平面推定時に、建物の境の奥行き情報を基に推定したため、建物に隣接する領域として描画されたものと思われる。

【3次元モデル統合処理】

ステレオペア 1,3,4 番により得られた入力画像ペアと、それにより獲得された3次元モデルをそれぞれ図 22, 23, 24に示す。また、これらの3次元モデルを統合した対象物体の生成結果を図 25(a)に示す。得られた統合結果は明らかに位置ずれが生じている。この原因として考えられることは、GPSと3軸角度センサで計測した各カメラ間の位置、姿勢データにノイズが含まれており、このノイズのために統合結果にずれが生じたものと考えられる。図 25(b)に手動で補正した統合結果を示す。

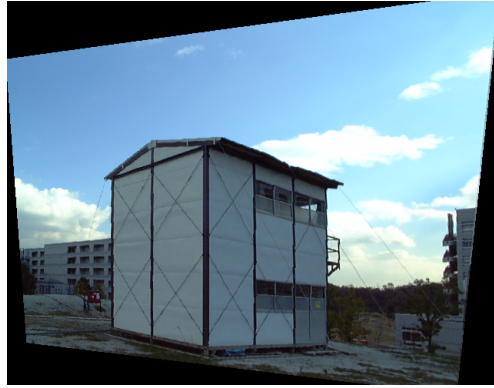
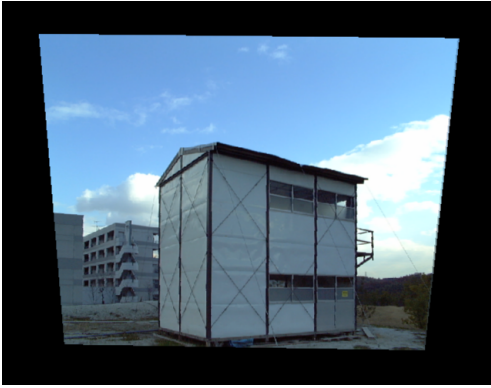
左画像 (基準画像)



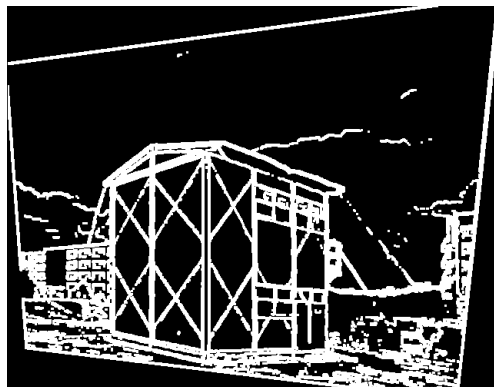
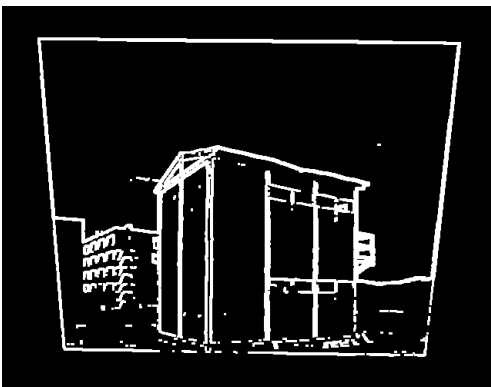
右画像 (参照画像)



(a) 入力画像



(b) エピポーラ線の平行化



(c) 特徴領域の抽出

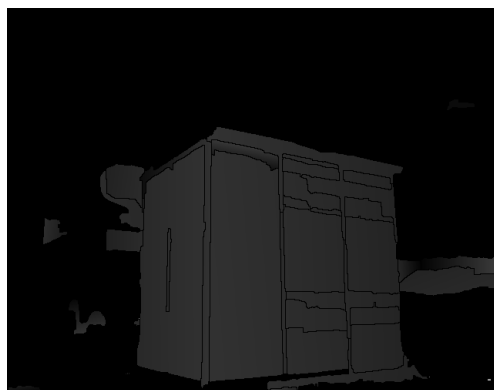
図 18 エピポーラ線平行化処理結果 (ステレオペア 2)



(a) 視差画像



(b) 基準画像に合わせた視差画像

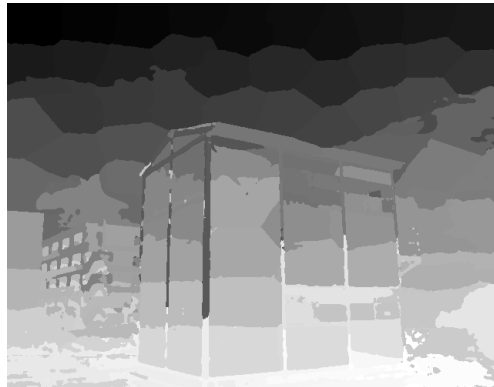


(c) 平面推定後の視差画像

図 19 視差推定処理結果 (ステレオペア 2)



(a) K 平均アルゴリズムによる分割結果

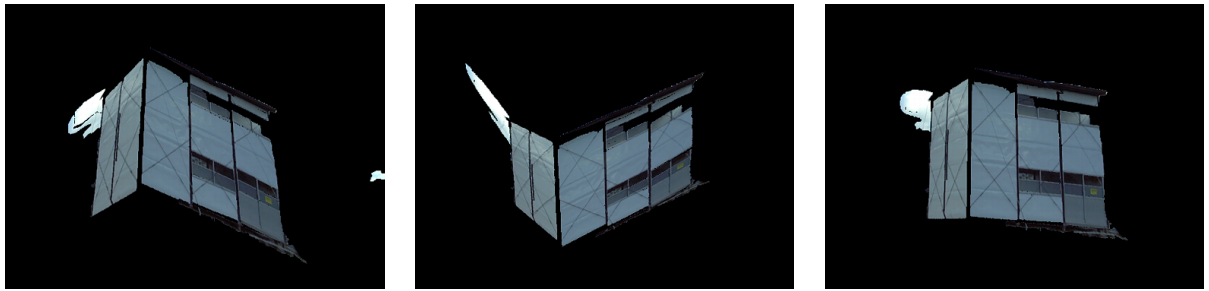


(b) 多数決フィルタによる処理結果

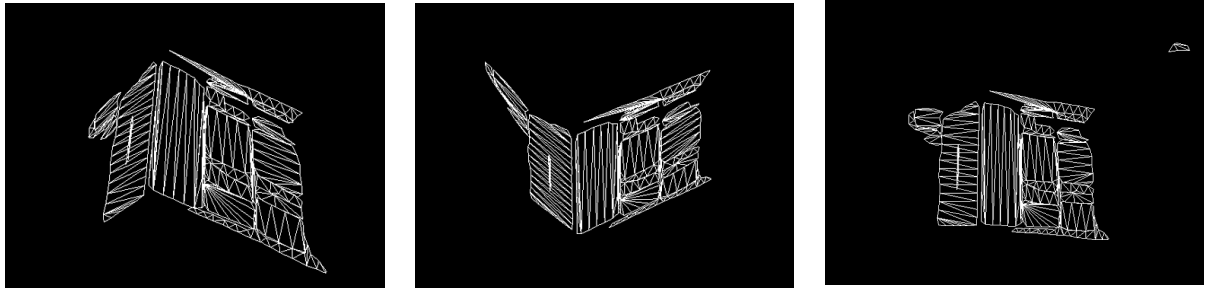


(c) クラスタ間の統合処理結果

図 20 領域分割結果 (図 18(a) の基準画像)



(a) 基準画像の 3 次元モデル



(b) ワイヤーフレームモデル

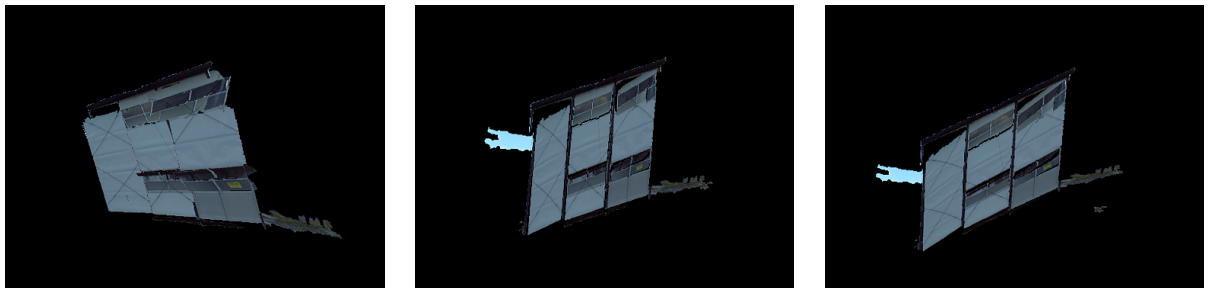
図 21 ステレオペア 2 による 3 次元モデルの獲得

基準画像

参照画像



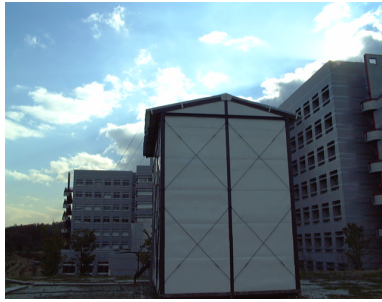
(a) 入力画像



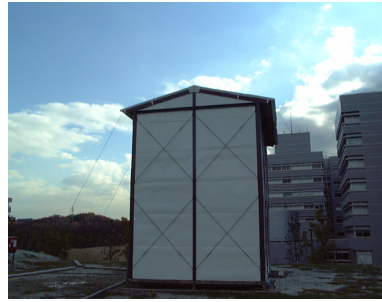
(b) 基準画像の 3 次元モデル

図 22 ステレオペア 1 による 3 次元モデルの獲得

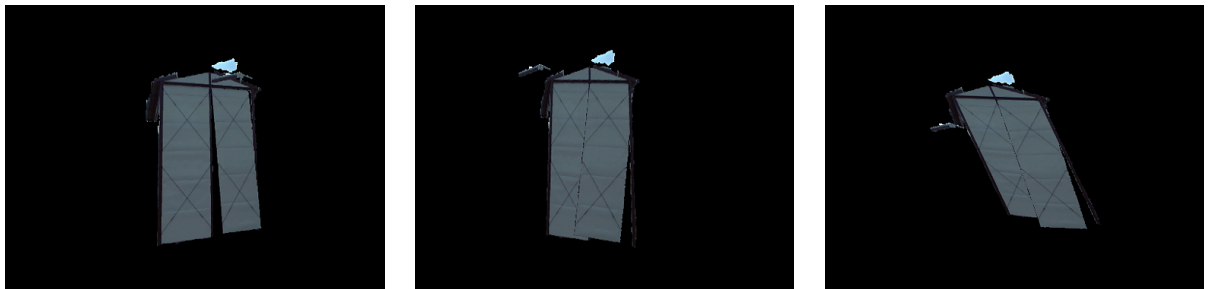
基準画像



参照画像



(a) 入力画像



(b) 基準画像の 3 次元モデル

図 23 ステレオペア 3 による 3 次元モデルの獲得

基準画像



参照画像

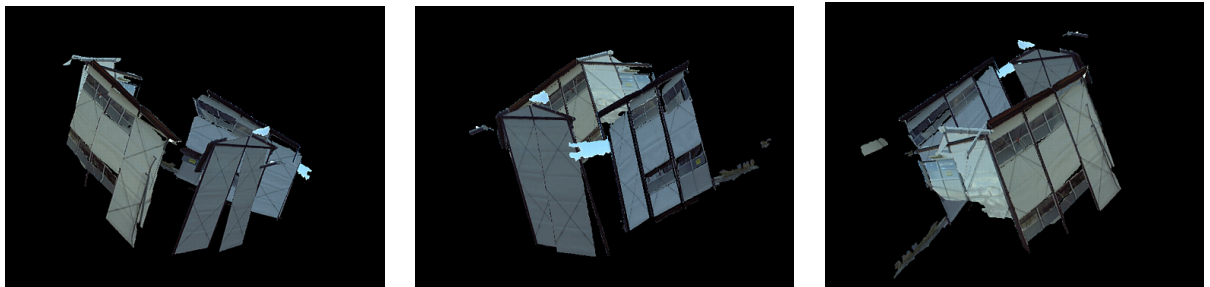


(a) 入力画像

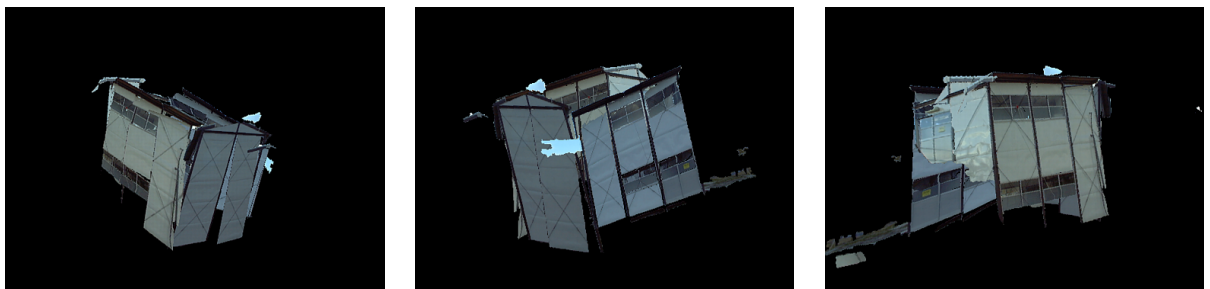


(b) 基準画像の 3 次元モデル

図 24 ステレオペア 4 による 3 次元モデルの獲得



(a) 3次元モデルの統合



(b) 補正後の3次元統合モデル

図 25 統合後の3次元モデル

5. 考察

本章では、屋外環境構造の再構成を行なった結果から、提案手法についての考察を行なう。

提案手法では、始めに、GPSと3軸角度センサにより撮影カメラ間の位置・姿勢の関係を求め、得られた計測結果を基にステレオ視を行ない奥行き情報を求めた。このように各センサを用いることにより、撮影場所や、照明条件に影響されずに、対象物体の撮影が行なえる。特に屋外環境下の場合、撮影場所に制限を受ける場合が多く、これらの各センサの使用は有効な手法であると言える。しかし、本研究で用いた各センサに関して幾つか問題点が挙げられる。まず3軸角度センサに関しては、計測時間に比例して計測誤差が蓄積していくため、計測時間が長ければ、時間に比例して計測結果に誤差が載ってしまう。また、GPSに関しては、計測時に使用される計測衛星の個数に応じて、測定精度が変化するという問題がある。図25(a)の結果から、統合後の各モデルの位置合わせに、明らかなずれも生じている。これは、各センサの誤差に起因するものであると考えられる。この誤差の問題を解決するために、得られた測定値の補正を行なう必要がある。この補正方法としては、ある対象物体が各画像上に重複して写っている部分の対応から、各カメラ間の位置・姿勢の関係を獲得して、得られた情報をもとに各センサの計測情報を補正することが考えられる。また、計測センサを増やすことによる、計測情報の冗長性から補正を行なうことも可能である。いずれにしても、各カメラの位置・姿勢情報の正確な計測、及び各計測情報の補正は本研究における重要な課題である。

次に、撮影画像に領域分割を施し、その結果とステレオ視により得られた奥行き情報から、ロバスト推定を用いて3次元平面の方程式を導出し、その撮影位置における3次元モデルを獲得した。このことは特徴領域でのステレオ視から得られる、疎な奥行き情報からでも、分割領域毎の平面方程式に沿った密な描画を行なうことができるという特徴をもつ。また、屋外環境下でのステレオ視の対応点探索問題に起因する誤対応の問題から、得られる奥行き情報全てが正しい値とは限らず、不正確な値も含まれている可能性が高い。そのような場合でもロバスト推定を平面方程式の導出時に適応することにより、不正確な値の影響を減らすこ

とができる。しかし、不正確な値が正しい値の半分以上を占めている場合には、ロバスト推定は適応できず、導出される平面の方程式もノイズの影響を受けた不正確なものとなってしまふ。この問題を回避するには、奥行き情報を獲得する時点において、いかに信頼性の高い奥行き情報を獲得するかという問題に帰着する。この問題の解決方法としては、現在 2 つの画像で対応点を探索している問題を、もう一つ参照画像を増やした 3 眼視のステレオを行なうことが考えられる。

最後に、各視点により得られた 3 次元モデルを統合する場合、現在は単に各センサにより得られた各撮影位置間の計測結果を基に統合しているだけであり、再構成結果も各センサの誤差値に直接影響されてしまふ。この問題に関しては、各視点において重複した部分を基に各モデル間の位置合わせを行ない、その位置合わせに基づいた、計測値の補正を行なうことなどが有効な解決策であると考えられる。

6. おわりに

本研究では屋外環境構造を仮想化して計算機内に再構成するために、多視点のステレオ画像から仮想空間を構築する手法を提案した。提案手法では、各撮影カメラ間の位置が厳密に固定された標準ステレオ撮像系を構成することは困難であると考え、複数視点からの撮影時に GPS と 3 軸角度センサを用いて、各撮影位置間の位置・姿勢の関係を獲得している。この関係からエピポーラ線の平行化を行ない、ステレオ法を用いて各々の視点における奥行き情報を獲得する。得られた多視点での奥行き情報を用いて作成する 3 次元モデルを統合して、最終的に屋外環境構造の再構成を行った。

提案手法の特徴として、各カメラ間の計測センサとして GPS と 3 軸角度センサを用いることにより、屋外環境下での撮影場所に影響されない対象物体の計測、及び撮影が行なえることが挙げられる。また、ステレオ視における対応点探索において、特徴領域上のみでの探索範囲で、その範囲内において輝度情報だけでなく色情報を積極的に用いた探索を行なうことにより、計算コストの削減、及び誤対応の影響を少なくした奥行き情報を獲得している。そして、得られた奥行き情報と、撮影画像に領域分割を施した分割結果から、領域毎の 3 次元平面をロバスト推定を用いて推定した。この処理により、ステレオ視の対応点探索の誤対応に起因する奥行き情報の誤値の影響も回避でき、特徴領域のみの疎な奥行き情報から、領域毎の密な奥行き情報が求まる。これらの処理手順を計算機上で実験し、提案手法による屋外環境構造の再構成を行なった結果を示した。

現状での問題点を幾つか挙げる。まず最初に、各カメラ間の位置、姿勢の関係を計測する各センサの計測値の誤差のため、再構成結果に著しく影響が及んでくる。そこで、この計測誤差を補正するための処理が必要となってくる。この補正方法としては、対象物体を多数の視点から撮影し、各画像上に重複して写っている部分の対応関係から、各カメラ間の位置、姿勢の関係を獲得して、各センサの計測情報を補正することが考えられる。また、ステレオ視における対応点探索の精度を上げることも重要である。ステレオ視を現在の 2 眼視から 3 眼視にするなど、精度の改善のために従来提案されているさまざまな手法を比較、検討し、より良い手法を提案する必要がある。

謝辞

本研究の全過程を通して、直接懇切なる御指導、御鞭撻を賜ったソフトウェア基礎講座 横矢 直和教授に衷心より感謝の意を表します。

本研究の遂行にあたり、終始有益な御助言と御鞭撻を頂いた像情報処理学講座 千原 國宏教授、並びに国内外の地より懇切丁寧に御指導頂いたソフトウェア基礎講座 竹村 治雄助教授に厚く御礼申し上げます。

さらに、本研究を通して、有益な御助言を頂いたソフトウェア基礎講座 岩佐 英彦助手、並びにソフトウェア基礎講座 山澤 一誠助手に厚く感謝します。

本研究を進めるにあたり、終始熱心に相談に応じて頂き、多くの時間と労力を割いて頂いたソフトウェア基礎講座の大隈 隆史氏、清川 清氏、並びに佐藤 哲氏に深く感謝します。

この2年間苦楽を共にし、有意義な研究生活、及び学生生活を共に過ごしてきたソフトウェア基礎講座の同輩諸氏、1年間ではありますが日々の研究室活動を様々な面で支えてくれた後輩諸氏に心より感謝致します。

最後に、研究室での生活を母のように見守って頂いたソフトウェア基礎講座事務補佐員 福永 博美女史、そして様々な面で支えてくれた諸先輩方、友人、そして両親に感謝します。

参考文献

- [1] G. G. Robertson, S. K. Card and J. D. Mackinlay: “Information Visualization Using 3D Interactive Animation,” *Communications of the ACM*, Vol. 36, No. 4, pp. 57–71, April 1993.
- [2] R. Teylingen, W. Ribarsky and C. Mast: “Virtual Data Visualizer,” *IEEE Trans. on Visualization and Computer Graphics*, Vol. 3, No. 1, pp. 65–74, March 1997.
- [3] S. Donikian: “VUEMS: A Virtual Urban Environment Modeling System,” *Proc. Computer Graphics International*, pp. 84–91, 1997.
- [4] O. Jokinen: “Building 3-D City Model from Multiple Unregistered Profile Maps,” *Proc. International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pp. 242–249, May 1997.
- [5] 全, 納富, 小澤: “動画像解析による仮想都市空間の構築”, *信学技報 PRMU96-126*, pp. 59–65, 1996.
- [6] O. Henricsson: “The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction,” *Computer Vision and Image Understanding*, Vol. 72, No. 2, pp. 163–184, November 1998.
- [7] 横光, 大隈, 竹村, 横矢: “多視点ステレオ実画像を用いた仮想空間の構築に関する一 検討”, *日本 VR 学会第 3 回大会大会論文集*, pp. 223–224, 1998.
- [8] A. Bensrhair, P. Miche and R. Debrie: “Fast and automatic stereo vision matching algorithm based on dynamic programming method,” *IAPR Pattern Recognition Letters*, Vol. 17, No. 5, pp. 457–466, May 1986.
- [9] C. Tomasi and T. Kanade: “The Factorization Method for the Recovery of Shape and Motion from Image Streams,” *Proc. Image Understanding Workshop*, pp. 459–472, 1992.

- [10] 佐賀, 安田, 横井: “実画像を利用した仮想空間の構築に関する研究”, Human Interface News and Report, Vol. 13, pp. 231–236, 1998.
- [11] P. E. Debevec, C. J. Taylor and J. Malik: “Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach,” Proc. SIGGRAPH '96, pp. 11–20, 1996.
- [12] 遠藤, 片山, 田村, 廣瀬, 渡辺, 谷川: “移動車両搭載カメラを用いた電脳映像都市空間の構築 (1)”, 日本 VR 学会第 2 回大会大会論文集, pp. 63–66, 1997.
- [13] 廣瀬, 渡辺, 谷川, 遠藤, 片山, 田村: “移動車両搭載カメラを用いた電脳映像都市空間の構築 (2)”, 日本 VR 学会第 2 回大会大会論文集, pp. 67–70, 1997.
- [14] A. Lippman: “Movie-Maps: An Application of the Optical Videodisc to Computer Graphics,” Computer Graphics, Vol. 14, No. 3, pp. 32–42, 1980.
- [15] S. E. Chen: “QuickTime VR - An Image-Based Approach to Virtual Environment Navigation,” Proc. SIGGRAPH '95, pp. 29–38, 1995.
- [16] M. Hirose, k. Yokoyama and S. Sato: “Transmission of Realistic Sensation: Development of a Virtual Dome,” Proc. Virtual Reality Annual International Symposium, pp. 125–131, 1993.
- [17] 廣瀬, 宮田, 谷川: “二次元実写画像を用いた三次元仮想空間の生成”, 信学技報 MVE96-18, 1996.
- [18] O. Grau: “A Scene Analysis System for the Generation of 3-D Models,” Proc. International Conference on Recent Advances in 3-D Digital Imaging and Modeling, pp. 221–228, May 1997.
- [19] K. Deguchi: “Factorization Method for Structure from Perspective Multi-View Images,” IEICE Trans. on Information and Systems, Vol. E81-D, No. 11, pp. 1281–1289, November 1998.

- [20] 杉本, 富田: “ステレオ画像の境界表現に基づくシーンの解釈”, 信学論 (D-II), Vol. J76-D-II No. 3, pp. 524-533, 1993.
- [21] 畑, 栄藤, 千原: “複数テンプレートによる多視点画像の記述”, 信学技報 PRMU96-130, pp. 9-16, 1997.
- [22] 徐: “CVCV-GW 特別報告 (XII): - Reconstruction の立場から: エピポーラ幾何, その復元と応用 -”, 情処学研報 CVIM103-8, pp. 59-66, 1997.
- [23] O. Faugeras: *Three-Dimensional Computer Vision : A Geometric Viewpoint*, The MIT Press, 1993.
- [24] 松山, 久野, 井宮: “コンピュータビジョン : 技術評論と将来展望”, 新技術コミュニケーションズ, 1998.
- [25] 大津: “テレビジョン画像情報工学ハンドブック”, オーム社, pp. 426-428, 1990.
- [26] 高木, 下田: “画像解析ハンドブック”, 東京大学出版会, pp. 485-491, 1993.
- [27] S. Mousset, P. Miche, A. Bensrhair and R. Debric: “Real-Time Determination of the Axial Motion Component from a Stereo Image Sequence,” Proc. International Conference on Image Processing and Its Applications, Vol. 4-6, No. 410, pp. 246-250, July 1995.
- [28] S. Z. Selim and M. A. Ismail: “K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality,” IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No. 1, pp. 81-87, January 1984.
- [29] P. J. Rousseeuw and A. M. Leroy: *Robust Regression and Outlier Detection*, John Wiley and Sons, NewYork, 1987.

- [30] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel: *Robust Statistics: An Approach Based on Influence Functions*, John Wiley and Sons, New York, 1986.

付録

A. カメラモデルと射影行列

本章ではカメラでの結像の仕組みを説明する。カメラでは前方に広がる実世界がレンズを通して2次元の世界、すなわち平面の画像へと投影され像を結ぶ。本章では、像の結び方としてカメラによる撮像過程の数理モデルである透視投影を扱い、それに付随する各カメラパラメータについて説明する。

A.1 透視投影モデル

実世界とレンズを通した投影面との位置関係を一番単純に理想化したモデルが、ピンホールカメラモデルである(図 26(a))。ピンホールカメラは結像の位置関係においては、そのレンズ中心にあたる位置にピンで穴を開けたスクリーンを置いたカメラと等価である。このピンホールの位置を光学中心という。実世界の点は、そこから光学中心を通る直線に沿って進んで、投影面と交わった所に像を結ぶ。

ピンホールカメラでは画像が逆さまにできるので、図 26(b)のように投影面を配置し直した方が分かりやすい。この配置では、光学中心から投影面にあたるスクリーンを透かして対象を見たように画像ができていると考えられるので透視投影と呼ばれる。

この透視投影モデルを数学的に定式化していく。図 27に示すような、光学中心に原点がある座標系を考える。Z 軸をレンズの光軸の方向にとる。これをカメラの光軸と呼ぶ。投影面は光軸と垂直になるように置かれるものとする。したがって、X-Y 平面が投影面と平行になる。このような座標系をカメラ座標系という。光学中心から投影面までの距離はレンズの焦点距離に相当するので、これを f とおく。するとこの座標系で $M(x, y, z)$ にある点の像は $m(f\frac{x}{z}, f\frac{y}{z}, f)$ にできる[23]。この中の投影面上への変換は式 (25) で表される。

$$(x, y, z) \longrightarrow (f\frac{x}{z}, f\frac{y}{z}) \quad (25)$$

この座標変換を透視変換と呼ぶ。左辺は空間の点のカメラ座標系での座標、右辺はその点の同じくカメラ座標上の位置である。

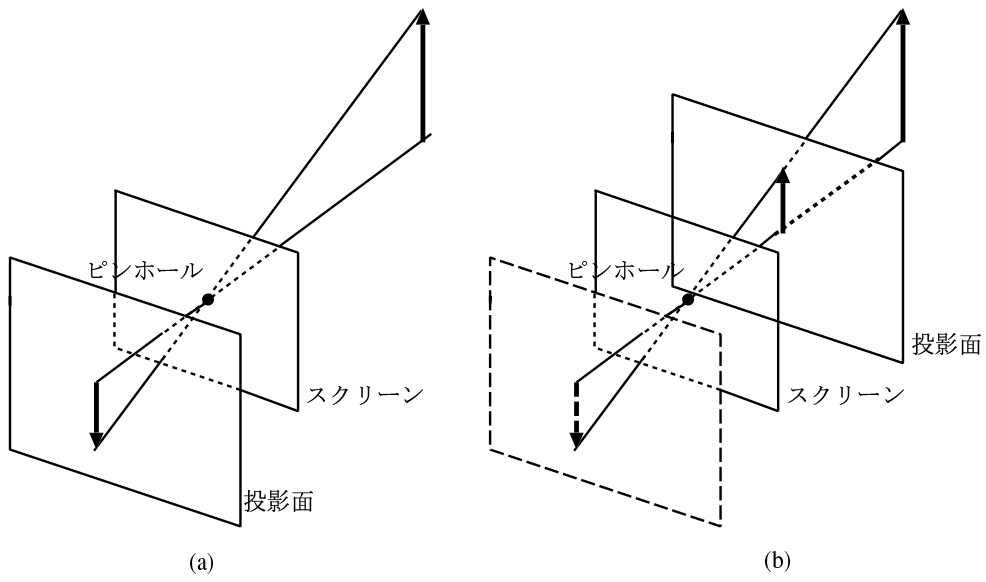


図 26 ピンホールカメラモデル

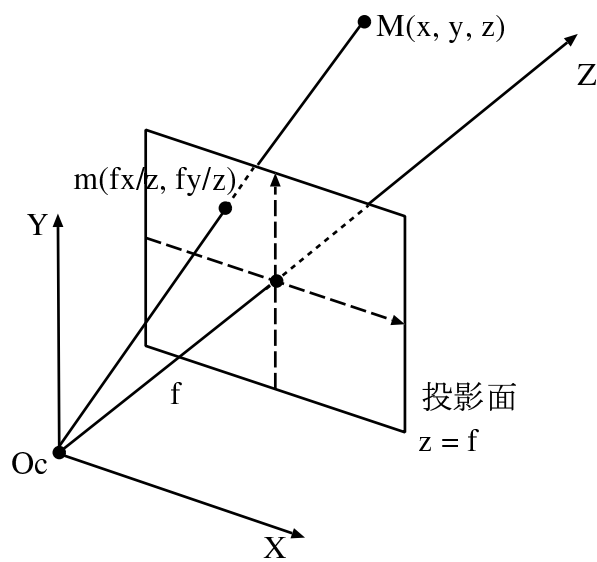


図 27 カメラ座標系と透視変換

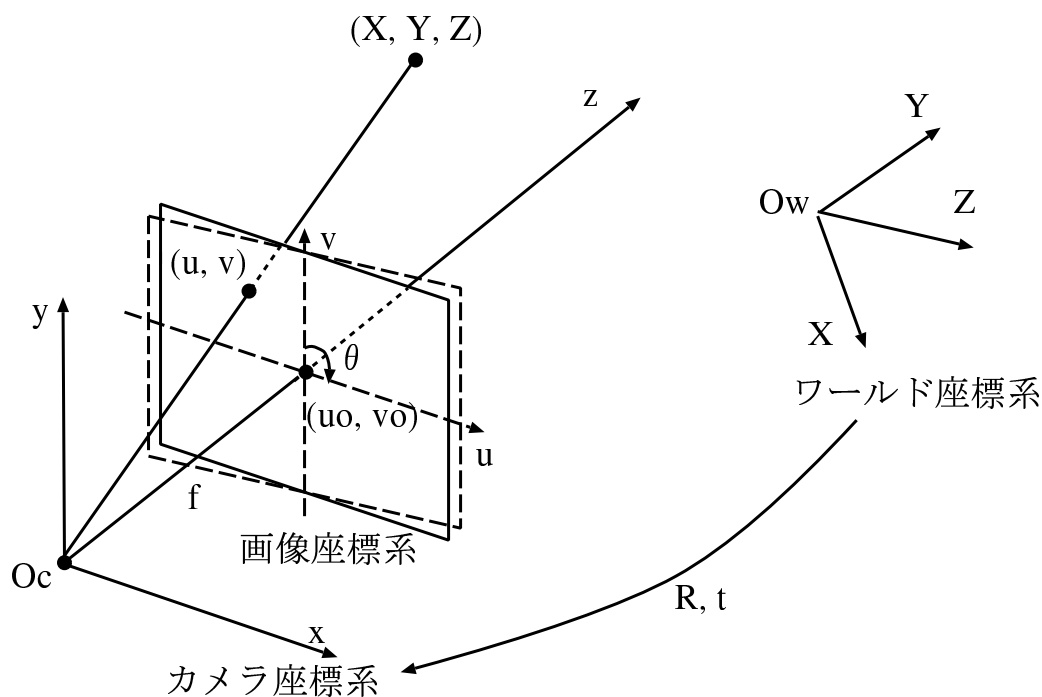


図 28 実際のカメラモデル

A.2 カメラパラメータ

本章では、カメラに関する幾何学的変換特性について説明していく。この幾何学的変換特性には大きく分けて、外部パラメータ、内部パラメータとあり、それぞれについて詳しく説明していく。

A.2.1 カメラ外部パラメータ

式(25)の透視変換は数式としては簡単なものであるが、ここで扱っているカメラ座標系は、カメラの光学系を単純にモデル化した上で設定したものである。実際のカメラでは、カメラの外形からはレンズの光学中心の位置、光軸の方向などは厳密に分からない。ここでカメラ座標系を独立に対象の位置、形状など「外部の世界」に基づいた座標系として図 28に示すように決め、すべてのカメラ座標系における光学中心の位置、光軸の方向などを統一的に扱える座標系を設定してお

く、このような独立な座標系をワールド座標系と呼ぶ。

ワールド座標系をカメラ座標系に変換するには、図 28 に示すように、各座標軸を回転させることで方向を一致させ、それから両座標系の原点がそろうまで平行移動する。座標軸の回転を 3×3 の行列、

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_x^T \\ \mathbf{r}_y^T \\ \mathbf{r}_z^T \end{bmatrix} \quad (26)$$

($\mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_z$ は、ワールド座標系でみたカメラ座標系の x, y, z 軸の方向を表す単位ベクトル) で表し、この平行移動ベクトル、すなわちワールド座標系の原点からカメラの光学中心へ向かうベクトルを、ワールド座標系で表すと \mathbf{T} 、一方、カメラ座標系でワールド座標系の原点を表すと \mathbf{t} であったとすると、ワールド座標 $\mathbf{X} = (X, Y, Z)^T$ で与えられる点はカメラ座標 $\mathbf{x} = (x, y, z)^T$ では、

$$\begin{aligned} \mathbf{x} &= \mathbf{R}\mathbf{X} - \mathbf{R}\mathbf{T} \\ &= \mathbf{R}\mathbf{X} + \mathbf{t} \end{aligned} \quad (27)$$

となる。したがって、カメラの位置と姿勢は行列 \mathbf{R} とベクトル \mathbf{T} 、または \mathbf{t} で表される。この回転行列 \mathbf{R} と並進ベクトル \mathbf{t} はカメラ外部パラメータと呼ばれる。

A.2.2 カメラ内部パラメータ

実際のカメラとピンホールカメラモデルの主要な違いを以下に挙げる。

1. 光軸すなわちカメラ座標系の z 軸は投影面の中心を通るわけではない。
つまり、光学的な画像の中心は投影面上での中心ではない。
2. 投影面は厳密には $x - y$ 平面と平行ではない。
3. 画像座標系でのスケールとカメラ座標系でのスケールが異なる。

これらの問題を考慮し、カメラ座標系と画像座標系の間では変換行列を求めていく。

図 28 に示すように、投影面上で光軸の横切る点 (画像中心) を (u_0, v_0) とする。次に、第 2 の問題として平行からのずれがわずかであれば、その効果は投影面上で直交する u, v 座標軸が空間的には直交していないことと等価であるとみなせる。そこで、投影面上の u 軸の方向にカメラ座標系の x 軸の方向をとるようにカメラ座標系を設定する。すると、カメラ座標系と画像座標系の違いは、原点の位置と、その縮尺と、 u, v 座標軸が空間的には直交していないことである。 u, v 座標軸が空間的になす角度を θ とすると、カメラ座標系で (x, y, z) の点は、透視変換によって投影面上

$$\begin{aligned} u &= f k_u \cdot \frac{x + y \cot \theta}{z} + u_0 \\ v &= f k_v \cdot \frac{y}{z \sin \theta} + v_0 \end{aligned} \quad (28)$$

に像を結ぶ。ここで f は焦点距離であり、 k_u, k_v は 1 画素の実質的な縦横の大きさの逆数である。この $k_u, k_v, \theta, u_0, v_0$ の変数はカメラ内部パラメータと呼ばれる。

A.3 同次座標系表示

図 28 に示した通り、空間の世界座標 $(X, Y, Z)^T$ と、画像座標 (u, v) との関係は式 (28) で表される。しかし、この式は数学的に扱いやすい形として以下の式でも表せる。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (29)$$

この式 (29) と式 (27), 式 (28) より

$$\mathbf{P} = \begin{bmatrix} f k_u & f k_u \cot \theta & u_0 \\ 0 & f k_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} (R \mid \mathbf{t}) \quad (30)$$

ここで、座標の表記に、

$$\begin{aligned}(X, Y, Z)^T &\rightarrow (X, Y, Z, 1)^T \\ (u, v)^T &\rightarrow (u, v, 1)^T\end{aligned}\tag{31}$$

とし、それぞれ 1 次元多いベクトルで表記している。この表記法を同次座標系表示と呼び、射影幾何学に由来するものである。この同次座標系を導入することで、特に座標間の関係を行列一つで表せる大きな利点がある。

この同次座標系を用いて、図 28 における空間上の点を $\mathbf{M} = (X, Y, Z, 1)^T$ 、その点の画像座標系での点を $\mathbf{m} = (u, v, 1)^T$ とすると、式 (29) は

$$\mathbf{m} = \mathbf{P}\mathbf{M}\tag{32}$$

である。この \mathbf{P} を射影行列といい、式 (32) を射影方程式という。この表記方法は 3.1 節で出てくるステレオ視を扱うときの重要な関係式になる。

B. ロバスト推定

ロバスト推定とは得られたデータの中に全体の傾向と全く異なった外れ値と呼ばれるデータが含まれている場合にこのような外れ値を除外したデータの傾向を検討するために用いられる手法である。ロバスト推定には以下のようなものがある。

- LMedS 推定

$$\text{Minimize}\{\text{med}(|r_i|)\}$$

残差 r_i (実測値と推定値の差) の絶対値の中央値が最小になるようなパラメータを最適なパラメータとして推定を行う手法。ただし残差の絶対値 $|r_i|$ の代わりに残差の 2 乗 r_i^2 が用いられることもある。一般に中央値は平均値よりも外れ値に対するロバスト性が高いため、中央値を最小化することで外れ値に対してロバストな推定結果を得ることが可能である。

- LTS 推定

$$\text{Minimize } \sum_{i=1}^h r_{i:n}^2$$

まず、残差 r_i を小さいものから順にソートし、ソート結果を $r_{i:n}$ とする。次にこのソート結果を 2 乗したものを h 個 (ただし h は $0 \leq h < n$ を満たす任意定数) 合計したものが最小となるパラメータを最適なパラメータとして推定を行う手法。外れ値は一般に残差が大きくなるため、残差の大きい値を除外して推定を行うことで外れ値に対してロバストな推定結果を得ることが可能である。

- M-推定

$$\text{Minimize } \sum_{i=1}^n \rho(r_i)$$

まず、 $\rho(-t) = \rho(t)$ を満たし、 $t = 0$ で唯一の最小値を持つ任意の関数 $\rho(x)$ を設定する。次に残差 r_i を関数 $\rho(x)$ へ代入し合計する。この合計が最小となるパラメータを最適なパラメータとして推定する手法。関数 $\rho(x)$ は統計処理の対象に応じて選択される。M-推定には最小 2 乗法も含まれるが、一般に M-推定という場合、より高次の多項式を利用した推定法を指す。一般に再帰的計算により推定を行うこととなり、初期値の設定が必要となる。