

Title	連続画像からの環境構造とカメラ運動の推定
Author(s)	陳, 謙
Citation	大阪大学, 1992, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3087961
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

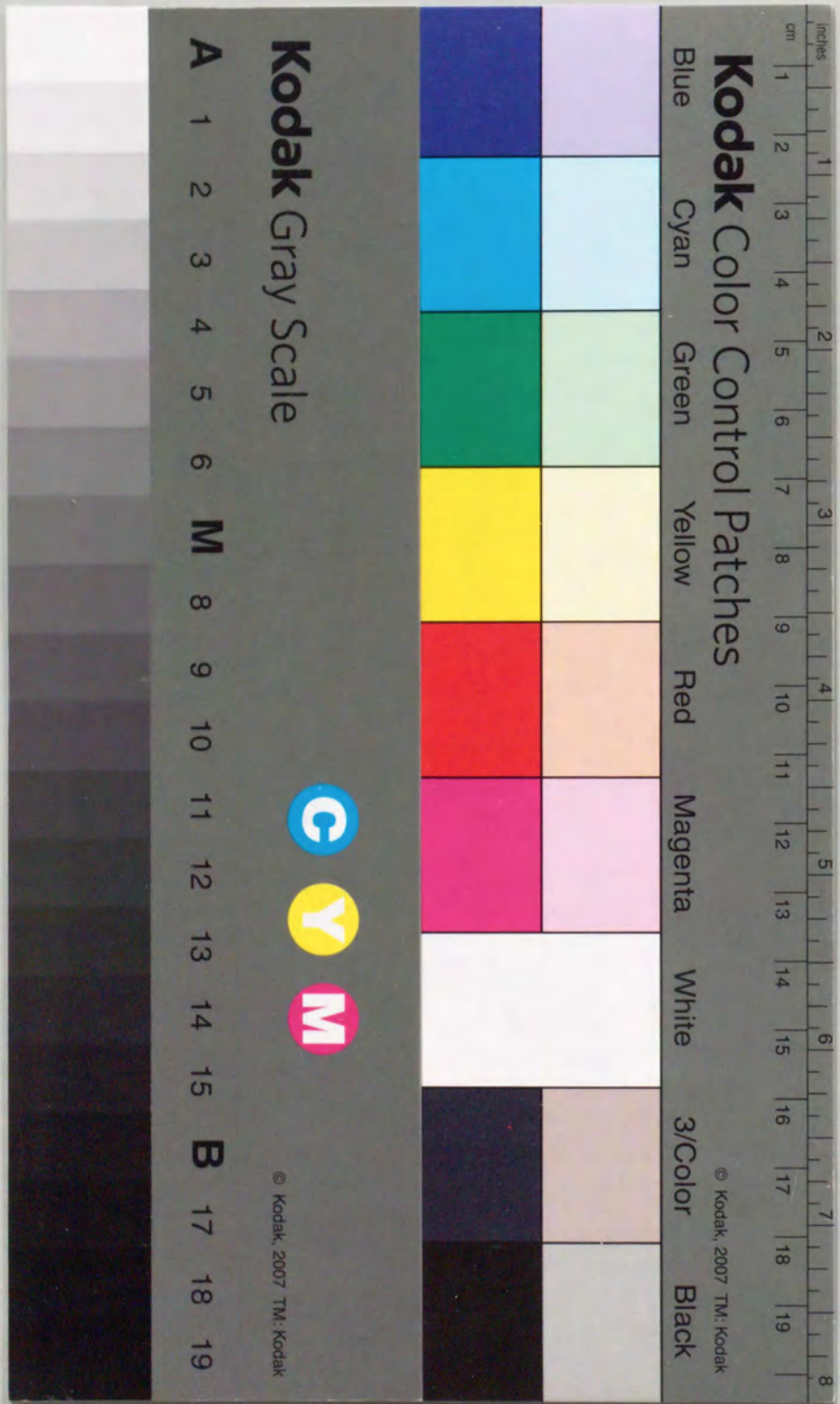
Osaka University

連続画像からの環境構造 とカメラ運動の推定

陳 謙

大阪大学基礎工学研究科

1992年1月



①

連続画像からの環境構造 とカメラ運動の推定

陳 謙

大阪大学基礎工学研究科

1992年1月

要旨

本論文は, 連続画像から環境構造及びカメラ運動の推定の研究について述べるものである. ロボットに“目”(カメラ)を持たせる目的は, 環境内の物体の形状や相対的位置関係を測定し, またロボット自身の動きを知る能力を持たせることである. この能力の実現を目的とする研究の中心課題は, カメラで撮像した画像から環境の3次元構造及びカメラ運動の推定である.

従来の研究はシーンから画像への投影を中心射影で表現し, 静止環境に対して画像内の特徴点の動きからカメラ運動と特徴点までの距離を算定する. しかし, 実画像に適用すると画像ノイズの影響によって算定された距離の誤差が大きい. そのため滑らかなカメラ運動という条件の下で多数回の観測結果を融合してノイズの影響を抑え, 推定の信頼性を向上させる研究がなされた.

しかし, この方法で環境内の物体形状を推定するには問題がある. 特にカメラから物体までの距離が遠いと形状推定は不正確になる. 画像の分解能に基づく距離算定の誤差は, 距離が遠くなると急速に増大する. 従って, 遠距離にある物体形状をこのような大きい誤差のある距離情報から算定しても信頼性のある結果を得ることは難しい.

物体までの距離が物体サイズに比べて十分大きいとシーンから画像への投影は平行射影で近似できる. 平行射影の利点の線形

要旨

性を利用し, 多数回の観測から物体形状とカメラ運動を信頼性高く推定する. しかし, 現実の世界での撮像条件が平行射影を満足するとは限らず, むしろ希であると考えられる.

一方, カメラの運動が制御可能の場合, ある拘束条件の下にカメラを動かし, カメラの運動情報を利用して物体の形状を抽出する研究がなされた. しかし, 現実には, 特に移動ロボットの場合, カメラの運動が自由に精密に制御できることは希であると考えられ, この方法の応用範囲は限られている.

本論文は, 一般化した撮像条件のもとで信頼性高く物体形状とカメラ運動を推定する方法を提案する. この方法の適用条件としては, (1) 環境にある動物体の数が少ないこと, (2) 動画像解析に伝統的に用いられたカメラ運動の平滑性, すなわちカメラ運動の並進・回転の両成分が急激には変わらないことを用いる.

本論文は, 一般的なシーンに対して形状推定をするため, 以下のアプローチで研究した.

1. 動物体の抽出
2. 静止物体の形状とカメラの運動の推定
 - a. 深さ範囲が狭いシーン
 - b. 一般のシーン

動物体の抽出はロボットの環境理解に有用な機能の一つである. 本研究では, 大多数の物体が静止しているという条件を利用して, 画像面上の動物体の像の動きと静止物体の像の動きの違い

要旨

から動物体を抽出する.

静止物体の形状とカメラ運動を求める時, まずシーンの深さ範囲が狭いと仮定する. この場合, シーンからカメラへの投影は平行射影で近似できる. カメラ運動が滑らかという条件から短い時間範囲での平行射影パラメータの変化は小さいと考えられる. 画像列を短いサブ画像列に分割すれば, 個々のサブ画像列におけるカメラの投影関係は一つの平行射影で近似できる. 平行射影の線形性を利用して物体形状を推定できる. 次に推定された形状は不変という条件を用いてサブ画像列間の平行射影のパラメータの変化, 即ち深さ方向の変化と回転角度の変化を個々のサブ画像列で求め, 物体形状とカメラ運動が推定できる.

深さ範囲が広い一般的シーンに本手法を適用すると, 得られた形状情報には歪がある. しかし, 特徴点の大まかな配置を知ることができる. この情報を利用して, シーンを分割し, ほぼ深さが一定のグループに分割する. 個々のグループに対して上記の手法を用いると, 各グループの形状とカメラ運動が計算できる. グループ間のパラメータの変化, 即ちスケールファクタの変化, 座標原点位置の変化と座標軸の回転角度の変化を, 算定されたカメラ運動は不変という条件を利用して求め, 各グループの形状をひとつに融合しシーン全体の形状を獲得することができる. このシーンの構造をカメラ運動を初期値として, ニュートン法を利用して投影の非線形方程式を解くことにより, 正確な環境構造とカメラ運動を

要旨

推定できる.

同じ方法を動物体に適用すると動物体の形状及び動物体とカメラ間の相対運動が求められる. この運動情報と静止物体に対するカメラの運動情報を利用して動物体の動きを算定できる.

目次

要旨	i
1 章 序論	1
2 章 視覚による環境構造の推定に関する研究	6
2.1 能動的センシングによる環境構造の推定	8
2.1.1 超音波センサによる距離の測定	8
2.1.2 投光法	9
2.1.3 レーザー光線を利用する方法	9
2.2 受動的センシングによる環境構造の推定	11
2.2.1 画像特徴に基づく研究	13
特徴点を利用する方法	13
ラインセグメントを利用する方法	15
動画像を利用する研究	16
2.2.2 オプティカルフローに基づく研究	19
2.3 能動的視覚	25
3 章 SVD による行列の分解及びノイズの除去	32
3.1 Singular Value Decompositon のアルゴリズム とその性質	33
3.2 SVD による行列の分解	35
3.3 SVD によるノイズの除去	37
4 章 連続画像を利用する動物体の抽出	44

目次

5章 深さ範囲が狭いシーンの構造の推定	52
5.1 行列の分解による形状情報と運動情報の分離	54
5.2 カメラの移動量が少ない場合での物体形状と カメラ運動の決定	59
5.3 カメラの移動量が多い場合の物体形状と カメラ運動の推定	65
6章 深さ範囲の広いシーンの構造の獲得	72
6.1 シーンの分割	75
6.2 分割されたシーンの融合による環境構造の復元	75
7章 ニュートン法によるシーンの構造とカメラ運動の推定	80
8章 実験と討論	84
9章 結論	104
謝辞	108
文献	110
研究業績	118

1 序論

ロボットは、色々な仕事を人間の代わりにしてくれる機械として開発された。今日、溶接ロボット、塗装ロボット、組立ロボット等の産業用ロボットは、自動車工場を始めとする様々の環境において活躍し、その威力を発揮している。しかし、より複雑で高度な仕事をする能力を持ち、多種多様な環境を柔軟に対応できるロボットを実現するには、機械工学、制御工学の分野の技術だけでは不十分であり、人間のような知能や感覚等の機能の実現が必要である。

視覚は人間及び大多数の動物にとって重要な情報源であり、人間が得る情報の90%以上は視覚を通じて獲得されると言われる。人間に近い能力を持つ知能ロボットを実現するために、視覚情報の利用は非常に重要である。近年、計算機を利用して人間或は動物の視覚の機能を実現しようとするコンピュータビジョン(マシンビジョン)の研究が盛んに行われ、注目を集めている。

人間や大多数の動物は視覚を通じて物体を認識し、その動きを推定し、3次元の環境内で移動できる。この機能はロボットにとっても必要であり、その実現には環境理解とロボット自身の位置と運動の推定、移動計画等の能力が必要である。視覚による環境理解の研究は1970年代から始められた。環境理解の一つの重要なテーマは、環境内の物体とロボットの間の相対的な位置関係及びロボット(カメラ)の運動情報の獲得、即ち、カメラで撮像した画像から環境の3次元構造とカメラ運動の推定の問題である。

従来の研究は、シーンから画像への投影を中心射影で表現し、静

止環境と仮定して画像内の特徴点の動きからカメラ運動と特徴点までの距離を算定する。しかし、実画像に適用すると画像ノイズの影響によって算定された距離の誤差が大きい。そのためカメラ運動が滑らかという条件の下で多数回の観測結果を融合しノイズの影響を抑え推定の信頼性を向上させる研究がなされた。

しかし、この方法で環境内の物体形状を推定するには問題がある。特にカメラから物体までの距離が遠いと形状推定は不正確になる。画像の分解能に基づく距離算定の誤差は、距離が遠くなると急速に増大する。従って、遠距離にある物体形状をこのような大きい誤差のある距離情報から算定しても信頼性のある結果を得ることは難しい。

物体までの距離が物体サイズに比べて十分大きいとシーンから画像への投影は平行射影で近似できる。平行射影の利点の線形性を利用し、多数回の観測から物体形状とカメラ運動を信頼性高く推定する。しかし、現実の世界での撮像条件が平行射影を満足するとは限らず、むしろ希であると考えられる。

最近、カメラの運動を能動的に制御することにより環境の理解の方法、いわゆる能動的視覚 (Active Vision) が提案された。カメラ運動が制御できる場合、ある拘束条件の下にカメラを動かす、その運動情報を利用して物体の形状を抽出する。その結果、Shape from X の問題で非線形問題が線形になったり、不安定な解法が安定になる等の利点があると報告された。

しかし、能動的視覚の手法を利用するには以下の条件が必要である。

1. カメラの正確な運動情報が獲得できる
2. カメラの運動が自由に制御できる
3. 適当な注視点を選択する

現実にはマニピュレータ等でカメラを動かす場合を除いて、正確な運動情報を知ることは困難で、カメラ運動は必ずしも自由に制御できない。更に、移動ロボットが働く環境には一般的に移動物体が存在する。本論文は、このような一般化した撮像条件のもとで信頼性高く物体形状とカメラ運動を推定する方法を提案する。この方法の適用条件としては、環境にある動物体は少数であることと、動画像解析に伝統的に用いてきたカメラ運動の平滑性、すなわちカメラ運動の並進・回転の両成分が急激には変わらないことを用いる。

本論文は、一般的なシーンに対して形状推定をするため、以下のステップで研究した。

1. 動物体の抽出
2. 静止物体の形状とカメラの運動の決定

動物体の抽出はロボットの環境理解に大きく役立つ機能の一つである。従来の研究は、静止環境或いは剛体運動を前提とする。しかし、一般的な環境には動物体が存在し、画像には異なる運動をする物体が存在する場合が多い。本研究では動物体の数が静止物体

に比べて少数である条件を利用して、動物体は存在しないと仮定して画像列を再構成する。再構成された画像上の特徴点と元の画像列上の特徴点の位置のずれの分析によって動物体を抽出する。

静止物体の形状とカメラ運動を求める時、まず物体の大きさとカメラ間の距離との比率(この後シーンの深さ範囲と呼ぶ)が小さいと仮定する。この場合シーンからカメラへの投影は平行射影で近似できる。カメラ運動は滑らかという条件から短い時間範囲での平行射影パラメータの変化は小さいと考えられる。画像列を短いサブ画像列に分割すれば、個々のサブ画像列におけるカメラの投影関係は一つの平行射影で近似できる。平行射影の線形性を利用して物体形状を推定する。次に推定された形状は不変という条件を用いて、サブ画像列間の平行射影のパラメータの変化、即ち深さ方向の変化と回転角度の変化を求め、物体形状とカメラ運動を推定する。

深さ範囲が広い一般のシーンに本手法を適用すると、得られた形状情報には歪がある。しかし、特徴点の大まかな配置を知ることができる。この情報を利用してシーンを分割し、深さがほぼ一定のグループに分類する。個々のグループに対して上記の手法を適用すると、各グループの形状とカメラ運動が計算できる。カメラ運動は不変という条件を利用して各グループの形状情報の参照座標系間のパラメータの変化、即ちスケールファクタの変化、座標原点位置の変化と座標軸の方向の変化を算定された求め、各グループの

形状をひとつに融合しシーン全体の構造を復元する。このシーンの構造をカメラ運動を初期値として、ニュートン法を利用して投影の非線形方程式を解くことにより、正確な環境構造とカメラ運動を推定する。

動物体が一つ以上ある場合、上記の方法を用いて各動物体の形状とカメラに対する運動を推定する。この運動情報と静止物体に対するカメラの運動情報を利用して動物体の動きが算定できる。

2 視覚による環境構造の 推定に関する研究

人間が得る情報の90%以上は視覚を通じて獲得されると言われるように、視覚は人間にとって重要な情報源である。人間及び大多数の動物は視覚を通じて物体を認識し、その動きを推定し、3次元の環境内で移動できる。人間のような能力を持つロボット、いわゆる知能ロボットの実現には、カメラで撮像される画像から環境内の物体の位置とロボット自身の運動を測定する機能が必要である。テレビカメラやコンピュータ等の機械を利用して、人間の視覚の機能を実現しようとする試みは1960年代から研究し始め、そして視覚を利用する環境構造の推定の研究は1970年代から始められた。

視覚による環境構造の推定に関する研究は利用されるセンサの種類によって能動的センシング (Active Sensing) と受動的センシング (Passive Sensing) に分類できる。能動的センシングとは超音波、光などのエネルギーを放出し、対象物体からの反射を分析によって、ロボットと物体の間の距離等を測定する手段である。受動的センシングとは受動的なセンサ (視覚の研究の場合、テレビカメラが良く使用される) を利用して撮像した画像を分析することにより環境構造を決定し、ロボット自身の運動を算定する方法である。この章では視覚による環境構造の推定に関する研究を回顧し、これまでの研究の流れ及び問題点について考査する。

2.1 能動的センシングによる環境構造の推定

環境構造の推定の基礎となる技術の一つはロボットと環境内の物体間の距離の測定である。能動的センシングの場合、超音波、光、電磁波等を発生し、環境内の物体に当て、テレビカメラ、アンテナ等のセンサで集められた反射信号を解析することによって物体の距離や形状を測定する。レーダは能動的センシングの代表的な例である。移動ロボットによく利用されるのは超音波センサによる方法、光のパターンを投光する方法とレーザー光線を利用する方法である。この節では能動的センシングの基本原則を紹介し、その特徴及び問題点について考査する。

2.1.1 超音波センサを利用する方法

超音波センサは超音波を発生し、環境内の物体に当てて、その反射を受信して、超音波の発生時刻と反射信号の受信時刻との間の間隔を測定することにより物体とロボットとの間の距離を計算する。

超音波センサの特徴として

1. 解像度が低い
2. 距離測定の精度が低い
3. 安価

等が挙げられる。このため、超音波センサは平面の多い屋内環境に働く移動速度の遅い自動運搬車のような屋内移動ロボットの衝突防止のに良く利用される。

2.1.2 投光法

この種の方法は特殊なパターンを光を物体に投光し、テレビカメラで撮像する。光のパターンは既知なので、画像から容易に認識できる。光源の位置、方向及びカメラの方向と位置を知れば、三角測定の原理を利用して物体とカメラの間の距離が測定できる。投光されるパターンの生成する方法としてプロジェクタを利用するものとレーザー光線を利用するものが挙げられる。

プロジェクタを利用する方法は多種類が存在するが、その共通の特徴は数枚の画像を撮像し、分析することによって広い視野範囲の距離情報が得られることである。一方、レーザーは点光源なので、広い視野範囲の距離情報を得るために、レーザー光線で測定される対象物体を走査する必要があり、大量の画像を撮像し、処理する必要がある。しかし、レーザー光線は非常に明るいので、画像からレーザー光線が当てた位置の抽出は容易である。

投光法には以下の特徴がある。

1. 物体に投光するので、物体表面の色、環境の照明条件等に影響されやすい。
2. 遠い物体の距離の測定は困難。
3. 獲得できる距離情報の密度はカメラの解像度以外に投光されるパターンにも制限される。

2.1.3 レーザー光線を利用する方法

レーザー光線を利用して距離を測定する手段として、三角測定

原理を利用するもの以外, レーダ, 超音波センサと同じように, レーザー光線を物体に当てて, 帰ってくる時間を測定して, 物体とロボットの間の距離を計算する.

この方法は投光法に比べて以下の特徴がある.

1. 距離測定の精度は物体との距離に影響されない.
2. 近い物体の距離測定の精度は投光法より劣る.

能動的センシングはセンサから超音波や光線などを放出して, その反射を利用して物体とロボットの間の距離を測定するので, 夜のような光のない環境においても機能する. しかし, その反面, 以下のような問題点がある.

1. 距離情報以外の物体の特徴の獲得は困難.
2. 物体が遠くなると反射信号は弱くなるので, 遠い物体の距離の測定は難しい.
3. 超音波, レーザー光線などを発生するので, 人体や環境内の他の物体に悪影響を与える恐れがあり, 安全上には問題がある.

それ以外, 超音波センサを利用する方法は密で精度の良い距離情報が得られなくて, 投光法の有効性は対象物体の表面の色などに影響されやすく, そしてレーザー光線を利用する方法には処理時間が時間が掛かり密な距離情報が得られない等の問題点がある.

2.2 受動的センシングによる環境構造の推定

静止環境内でカメラを移動させると(回転運動と並進運動のどちらも含む), カメラの運動に伴って, 画面上の物体の投影も運動する. この運動は画像上の特徴或いは明るさのパターンの運動を観測することによって抽出でき, この運動情報を分析することにより環境構造とカメラ運動を推定することができる. 環境構造と同様にカメラ運動は非常に重要な情報である. カメラ運動の情報を利用しなくてもステレオ視等の手法で環境構造が決定できる. しかし, カメラの運動は分かれば画像のセグメンテーションが容易にでき, ステレオ視の対応付け問題も容易に解決できる.

従来の研究は利用する画像特徴の違いによって2種類に分かれる. 一つはコーナや Occluding Boundary のような物体の3次元形状特徴に対応する画像特徴に基づく研究である. この種の研究は, 先ず画像からエッジ点, ラインセグメント等の特徴を抽出し, これらの画像特徴の画像間の対応付けを行い, そして剛体運動, 即ち剛体上の二つの点の間の距離は不変という拘束条件を利用して, 画像特徴の位置変化とシーンの中の物体の形状, 位置及びカメラ運動のパラメータとの間の関係を表す方程式を書き, これらの方程式を解くことによりシーンの構造とカメラ運動のパラメータを求める.

もう一つはオプティカルフローに基づく研究である. この方法は, まず画像列から画像上の画素の移動速度(オプティカルフロー

と呼ぶ)を抽出する。求められたオプティカルフローと他の拘束条件を利用してシーンの中の物体とカメラ間の3次元運動速度を求める。

従来の方法はカメラを移動ロボット本体に固定して、ロボットが移動しながら撮像し、カメラから得られる連続な画像列を利用して環境の構造及びカメラ運動を抽出する。この場合、カメラはロボットの本体に固定されているので、カメラ運動は自由に制御できない。ロボットの車輪に付いているエンコーダ等から得られるロボットの運動情報の誤差が大きく、カメラのカレブレーションは困難等の原因で、環境構造の決定に利用できるカメラ運動の情報はほとんどないと考えられる。このようにカメラの運動の制御を行わず、カメラの運動情報を利用しない方法は受動的視覚とも呼ばれる。

近年、人間が物を見る行為に注目し、注視制御というアイデアをコンピュータビジョンの研究に取り入れ、能動的視覚(ACTIVE VISION)と呼ばれる新しいパラダイムが提案された。注視制御を行うことにより、受動的視覚に比べて物体中心の座標系で表現する物体の形状情報は容易に獲得でき、不良設定(ILL-POSED)問題は良設定(WELL-POSED)になる等の利点があると報告された。

この節では今までの研究の流れを回顧し、この分野の研究の進展を把握するとともに、従来の研究の特徴及び問題点について考

査する。

2.2.1 画像特徴に基づく研究

画像特徴を利用する方法は先ず画像列からエッジ点、ラインセグメント等の画像特徴を抽出し、これらの画像特徴に対して画像間の対応付けを行い、そして剛体運動、つまり剛体上の二つの点の間の距離は不変という拘束条件を利用して、画像特徴の位置変化とシーンの中の物体の形状、位置及びカメラ運動のパラメータとの間の関係を表す方程式を書き、これらの方程式を解くことによりシーンの構造とカメラ運動のパラメータを求める。

画像特徴の抽出及び画像間における画像特徴の対応付けについては画像処理、2眼ステレオ視等の研究の中で長く研究されてきた。この節では画像特徴は既に抽出され、これらの画像特徴の画像間の対応付けは既に求められていることを前提とする画像特徴の動きからシーンの中の物体の形状、位置及びカメラ運動の推定に関する研究を紹介する。

特徴点を利用する研究

画像特徴の動きを利用して環境構造とカメラ運動を決定する可能性について何人かの研究者によって解明された。Ullman^{[32][33]}はシーンからカメラへの投影は平行射影であることを前提にして、画像特徴の動きから環境構造とカメラ運動の推定について研究し、空間内の4つの特徴点を3回観測することによってカメラ

運動と特徴点の位置を決定できることを示した。

しかし、カメラの投影は中心射影であり、少数の場合を除いて平行射影で近似できない。中心射影の場合、3次元の空間座標系から画像面の2次元座標系への投影変換は非線形なので、画像特徴の動きから環境構造とカメラ運動を決定する問題は平行射影の場合に比べて複雑になる。Roach と Aggarwal^{[24][25]}は、静止環境においてカメラが運動し、シーンからカメラへの投影は中心射影である場合、空間内の点の位置とカメラの運動を決定する可能性について研究した。ある特徴点の3次元座標 (X, Y, Z) とその点の像の座標の間の関係は次の式で表現できる。

$$\begin{cases} x = F \frac{a_{11}(X - X_0) + a_{12}(Y - Y_0) + a_{13}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)} \\ y = F \frac{a_{21}(X - X_0) + a_{22}(Y - Y_0) + a_{23}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)} \end{cases}$$

ここで、 F は焦点距離、 (X_0, Y_0, Z_0) は投影中心、 $a_{11}, a_{12}, \dots, a_{33}$ はカメラの光軸の方向 (Θ, Φ, Ψ) の関数である。これらのパラメータを決定するには、2枚の画像上で対応付けが取れた5対の画像特徴点が必要である。Roach と Aggarwal は Levenberg-Marquardt のアルゴリズムを用いてカメラの投影方程式を解き、空間内の点の位置とカメラ運動を推定する方法を提案した。この問題に関して Nagel^[19] はカメラの平行移動と回転は独立に求められる解方を提案した。

Tsai と Huang^{[30][31][13]} はシーンの特徴がある平面上に分布している場合の問題について研究した。彼らは2つの基本ステップ

を分けてシーンの構造とカメラ運動を推定する。まず8個の“純パラメータ”を定義し、2枚の画像を利用してこれらの“純パラメータ”を未知数とする線形方程式を書き、これらの方程式を解くことにより“純パラメータ”を求める。カメラの運動パラメータはこれらの“純パラメータ”から計算できる。

カメラの投影関係及び剛体運動の拘束条件を分析して、平面 $AX + BY + CZ = 1$ 上の特徴点の2枚の画面上の像の座標間の関係は次の式で表現できる。

$$\begin{cases} x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \\ y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \end{cases}$$

ここで、 (x, y) と (x', y') はそれぞれ特徴点の1枚目の画像上の座標と2枚目の画像上の座標であり、 a_1, a_2, \dots, a_8 は“純パラメータ”で、カメラの焦点距離、構造パラメータ (A, B, C) 、カメラの並進運動と回転運動のパラメータで表すことができる。構造パラメータと運動パラメータは“純パラメータ”から算定できる。

Tsai と Huang^[31] 及び Longuet-Higgins^{[17][18]} はこの方法を発展させ、特徴点が曲面の上に分布する場合の問題を研究し、他に何人かの研究者^{[42][43][44][6][10]} はこの研究を拡張した。

ラインセグメントを利用する研究

特徴点を利用する研究以外、ラインセグメントを利用して環境構造とカメラ運動を推定する研究もなされた。3次元直線の自由

度は3次元特徴点に比べて1つ多いので、環境構造とカメラ運動を決定するための最低限の画像のフレーム数は特徴点を利用する場合に比べて多くなる。

Yen と Huang^{[38][39]}は球面投影に基づく繰り返しアルゴリズムを提案した。この方法を利用して環境構造を決定するには最低7本の直線と3枚の画像が必要である。

Liu と Huang^{[15][16]}は、まずカメラの回転運動を推定し、そしてカメラの並進運動を推定する。回転運動のみ存在する場合、回転パラメータを決定するには2枚の画像間に対応付けを取れた2本の直線が必要で、並進運動のみ存在する場合、3枚の画像間に対応付けを取れた5本の直線が必要であり、一般的な場合、3枚の画像と6本の直線が必要である。この方法はノイズと繰り返し解法の初期値に非常に敏感で、そして並進運動のパラメータは回転パラメータの誤差に強く影響される。

その他、Faugeras, Lustman と Torscani^[6]はカルマンフィルタを導入してノイズの影響に強い方法を提案した。

動画像を利用する研究

少数枚の画像を利用して環境構造とカメラ運動を推定する研究以外、カメラから連続に撮像された画像列を利用して環境構造とカメラ運動を推定する研究も行われた。Ullman^[34]は平行射影に基づく画像列からシーンの構造を決定する incremental な方法を提案した。物体の形状を行列 $M(t)$ で表現する。行列 $M(t)$ の要

素はカメラ中心の座標系で表す t 時刻の物体の特徴点の3次元座標 (X_i, Y_i, Z_i) である。カメラへの投影は平行射影なので、 (X_i, Z_i) は特徴点の画面上の像の座標と同じである。従って、環境構造の決定は Y_i を推定することになる。剛体運動と仮定して、 (X'_i, Y'_i, Z'_i) は t' 時刻での点 i の位置とする。 Y'_i を剛体運動から離れる傾向が最小になるべきという条件を利用して求める。 L_{ij} と L'_{ij} はそれぞれ t 時刻と t' 時刻の点 i と点 j の間の距離とし、剛体運動を離れる傾向を次のように定義する。

$$E = \sum_{I,J} \frac{(L_{ij} - L'_{ij})^2}{L_{ij}^3}$$

この方法は最初の数フレームの画像から3次元構造の良い近似が得られるが、真の値には収束しない。

Incremental な方法の主な欠点は物体の回転軸が固定している以外の場合には良好な結果は得られないことである。そして、平行射影の仮定をしているので利用できる範囲は更に狭くなる。

Broida と Chellappa^[45]はシーンには1つの剛体だけが存在し、この物体の並進運動の速度と回転速度は変化しない場合の問題に対して、カルマンフィルタを利用する解法を提案した。

Tomasi^[28]らは、距離算定をしないで遠距離にある物体の形状を正確に推定する方法を提案した。物体までの距離が物体サイズに比べて十分大きいとシーンから画像への投影は平行射影で近似でき、平行射影の利点の線形性を利用し、多数回の観測から物体

形状とカメラ運動を信頼性高く推定する. カメラ座標系は図1が示すよう設定し, 物体の特徴点の平均距離は物体の大きさに比べると十分大きい場合, カメラの投影関係は次の式で近似できる.

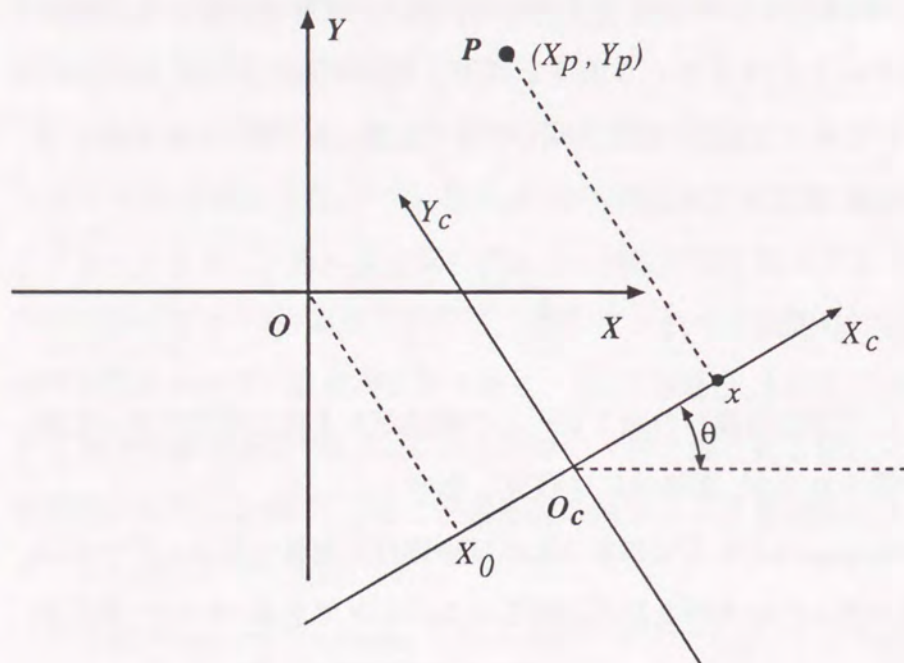


Figure 1 The camera coordinate system.

$$x = X_p \cos \theta + Y_p \sin \theta + X_0$$

ここで x は点 P の像の座標で, (X_p, Y_p) は点 P の位置で, θ はカメラの光軸と世界座標系の Y 軸の間の角度で, X_0 はカメラ座標系での世界座標系の原点の X 座標である.

Tomasi らは, 画像列とカメラの運動パラメータ及び物体の特徴点の位置の間の関係は次の式で表現する.

$$I = MS$$

ここで

$$I = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$$

$$M = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 & X_{01} \\ \cos \theta_2 & \sin \theta_2 & X_{02} \\ \vdots & \vdots & \vdots \\ \cos \theta_m & \sin \theta_m & X_{0m} \end{pmatrix}$$

$$S = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

線形代数の Singular Value Decomposition のアルゴリズムを利用して行列を分解し, この方程式を解けた. この方法を実画像に適用して, 非常に良い結果が得られたと報告された. この手法の特徴のひとつは, 通常の方法のようにカメラ運動の平滑性は必要なく, 平行射影の条件を満たせば良いことである.

2.2.2 オプティカルフローに基づく研究

カメラ或いは環境内の物体の運動に伴い, 画面上の物体の投影も運動する. オプティカルフローとは動画像上の明るさのパターンの移動速度である. オプティカルフローと他の拘束条件を一緒

に利用してシーンの中の物体とカメラ間の3次元運動速度を求めることができる。

この方法は画像特徴を利用する方法に比べて以下の特徴がある。

1. オプティカルフローは画像特徴より密なので、シーンの密な構造情報が得られる。
2. オプティカルフローは画像の明るさの時間及び空間に対する微分で求められるので、画像間の対応付けを行う必要はない。

オプティカルフローによる環境構造の推定は二つの基本ステップに分かれる。

1. 画像の明るさの変化からオプティカルフローを抽出する。
2. オプティカルフローを利用して環境構造とカメラ運動を推定計算する。

ここでは1に関する研究の紹介を省略し、2に重点をおいて考查する。

図2が示すように、カメラ中心の座標系 $O-XYZ$ を利用する。座標原点をカメラの焦点に設定し、 Z 軸をカメラの光軸と平行するように設定する。議論を簡単にするためにカメラの焦点距離 f を1とする。画像面の座標系を $o-xy$ とし、 x 軸と y 軸をそれぞれ X 軸と Y 軸と平行するように設定する。

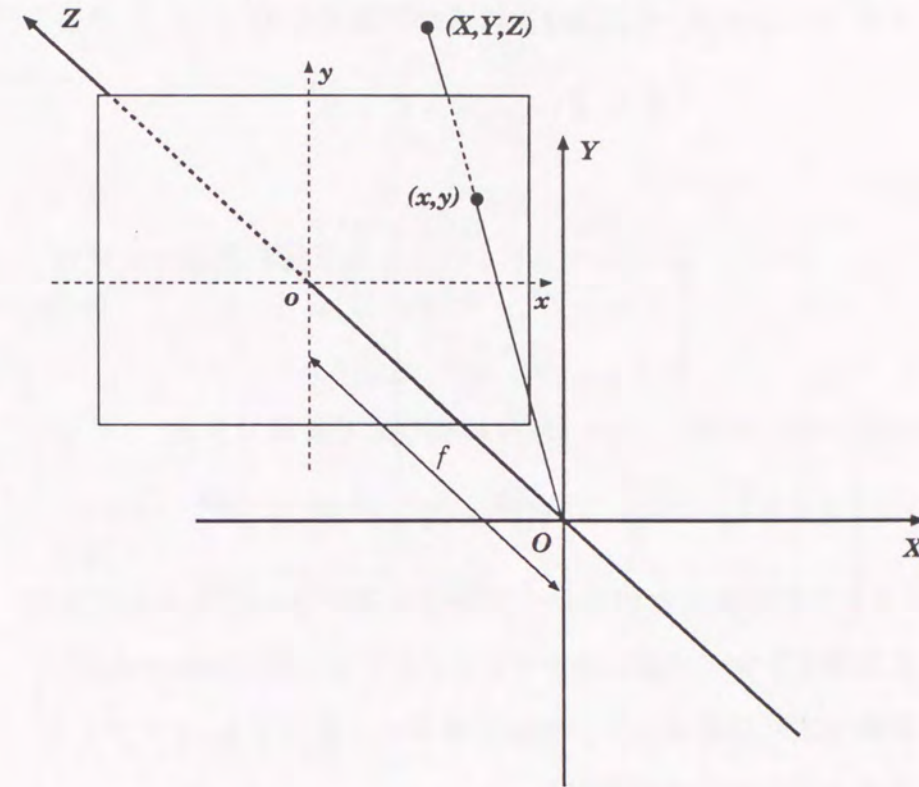


Figure 2 Camera centered coordinate system.

中心射影により、点 (X, Y, Z) の像の画像面上の座標は次の式で計算できる。

$$\begin{cases} x = \frac{X}{Z} \\ y = \frac{Y}{Z} \end{cases} \quad (2.1)$$

カメラの並進運動の速度と回転速度は

$$\begin{cases} \mathbf{V} = (V^X, V^Y, V^Z) \\ \boldsymbol{\Omega} = (\Omega^X, \Omega^Y, \Omega^Z) \end{cases}$$

とする. 点 (X, Y, Z) の速度は次の式で計算できる.

$$(\dot{X}, \dot{Y}, \dot{Z}) = -(\mathbf{V} + \boldsymbol{\Omega} \times \mathbf{R})$$

即ち

$$\begin{cases} \dot{X} = -V^X - \Omega^Y Z + \Omega^Z Y \\ \dot{Y} = -V^Y - \Omega^Z X + \Omega^X Z \\ \dot{Z} = -V^Z - \Omega^X Y + \Omega^Y X \end{cases} \quad (2.2)$$

その点の像の速度 $(u, v) = (\dot{x}, \dot{y})$ は次の式で計算できる.

$$\begin{cases} u = x \frac{V^Z}{Z} - \frac{V^X}{Z} + xy\Omega^X - (1-x^2)\Omega^Y + y\Omega^Z \\ v = y \frac{V^Z}{Z} - \frac{V^Y}{Z} + (1+y^2)\Omega^X - xy\Omega^Y - x\Omega^Z \end{cases} \quad (2.3)$$

環境構造とカメラ運動はオプティカルフロー及び物体の表面には急激な変化は存在しない仮定に基づいて推定する. オプティカルフローは次の式で近似でき,

$$\begin{cases} u(x, y) = u_0 + u_x x + u_y y + u_{xx} x^2 + u_{xy} xy + u_{yy} y^2 + O_3(x, y) \\ v(x, y) = v_0 + v_x x + v_y y + v_{xx} x^2 + v_{xy} xy + v_{yy} y^2 + O_3(x, y) \end{cases} \quad (2.4)$$

そして物体の表面は次の式で近似できる.

$$Z = Z_0 + Z_X X + Z_Y Y + \frac{1}{2} Z_{XX} X^2 + Z_{XY} XY + \frac{1}{2} Z_{YY} Y^2 + O_3(X, Y) \quad (2.5)$$

(2.1) と (2.5) から次の式が得られる.

$$Z(x, y) = \frac{Z_0}{1 - Z_X x - Z_Y y - \frac{1}{2} Z_{XX} x^2 - Z_{XY} xy - \frac{1}{2} Z_{YY} y^2 - O_3(x, y)} \quad (2.6)$$

ここで

$$Z_{xx} = Z_0 Z_{XX}$$

$$Z_{yy} = Z_0 Z_{YY}$$

$$Z_{xy} = Z_0 Z_{XY}$$

カメラの並進運動速度を次のように変換する.

$$\begin{cases} V^x = \frac{V^X}{Z_0} \\ V^y = \frac{V^Y}{Z_0} \\ V^z = \frac{V^Z}{Z_0} \end{cases} \quad (2.7)$$

剛体運動の仮定に基づき, 上記の式から次の方程式が得られる.

$$\begin{cases} u_0 = -V^x - \Omega^Y & v_0 = -V^y + \Omega^X \\ u_x = -V^z + V^x Z_X & v_x = -\Omega^Z + V^y Z_X \\ u_y = \Omega^Z + V^x Z_Y & v_y = V^z + V^y Z_Y \\ u_{xx} = -2V^z Z_X + V^x Z_{xx} - 2\Omega^Y & v_{xx} = V^y Z_{xx} \\ u_{xy} = -V^z Z_Y + V^x Z_{xy} + \Omega^X & v_{xy} = -V^z Z_X + V^y Z_{xy} - \Omega^Y \\ u_{yy} = V^x Z_{yy} & v_{yy} = -2V^z Z_Y + V^y Z_{yy} + 2\Omega^X \end{cases} \quad (2.8)$$

(2.8) はオプティカルフロー, オプティカルフローの1次, 2次微分, 物体の形状とカメラ運動の関係を表す方程式である. 物体の表面は Z_X, Z_Y, Z_{xx}, Z_{xy} と Z_{yy} で表現でき, 運動パラメータは \mathbf{V} と $\boldsymbol{\Omega}$ で表現できる. 従って, 物体の形状とカメラ運動はこの12個の非線形方程式を解くことにより求められる. 何人かの研究者

はオプティカルフローから環境構造とカメラ運動の推定についての研究を行なった^{[26][46][35][36][22][1][37][23][21][5][11][12][20][8]}.

画像列から環境構造とカメラ運動の推定の問題について画像特徴を利用する方法とオプティカルフローを利用する方法が提案された. この2つの手法にはそれぞれの利点と欠点がある.

画像特徴を利用する方法は画像特徴の画像のフレーム間の対応付けが必要である. 対応付けの問題はステレオビジョン等の分野において長く研究されたにもかかわらず, 完全解決までは至っていない. オプティカルフローを利用する方法は画像間の対応付けを必要としない. しかし, オプティカルフローの計算, そしてそれを利用するシーンの構造とカメラ運動を推定するとき, 画像の明るさに対して1次微分及び2次微分を計算する必要がある. 微分演算はノイズを強調する効果があるので, 実画像にはたくさんのノイズが含まれており, 今までに提案されたオプティカルフローを利用して環境構造とカメラ運動を推定する方法には実用的な結果をもたらすことができるものはほとんどなかった.

画像特徴に基づく方法は画像ノイズに対してオプティカルフローに基づく方法ほど敏感ではないが, 得られた結果が画像ノイズに強く影響されることは変わらない. ノイズの影響を抑えるために, 環境の構造を決定できるのに最低限の特徴点の数と画像の枚数より多くの特徴と画像を利用し, 多数回の観測結果を融合して信頼性を高めようとする研究は行われた. 例えば, Tomasiら

は数百枚の画像を利用して, かなり信頼性の良い結果が得られたことを報告された.

2.3 能動的視覚

人間が物を見るとき, 関心のある物体の特徴を良く見るために, その物体に近づいたり, 離れたり, 視点を変えたりする. 人間のこのような普段の生活の中で自然に行う物を見る行為を注目して, 画像から抽出した情報を分析し, 物体の特徴が良くわかるようにカメラを能動的に制御する, いわゆる注視制御というアイデアをコンピュータビジョンの研究に取り入れ, 能動的視覚 (Active Vision) という新しいパラダイムが最近提案された. カメラの注視制御によって, 受動的視覚に比べて, 物体中心の座標系で表現する物体の形状情報が容易に獲得でき, 不良設定 (ill-posed) 問題は良設定 (well-posed) になる等の利点がある.

Aloimonos^[2]は Structure from Motion, Shape from Shading, Shape from Texture, Shape from Contour の問題に対してカメラの能動的制御を導入し, 観測される物体上の1つの特徴点を画面の中心に保ちながらその特徴点を円心とする円弧に沿って移動するようにカメラを制御することにより, 従来の方法に比べて問題は良設定 (well-posed) になったり, 非線形の問題は線形問題になったり, 不安定な解法が安定になる等の良い特徴が発見された.

石黒^[47]らはカメラを直線に沿って動しながら注視制御により環境の局所地図を獲得する問題について研究した. カメラを精密

に回転させることができる回転台に載せ、移動ロボットに設置する。ロボットが直線に沿って一定な速度で移動しながらカメラを環境内の一本の垂直エッジを注視させ、その垂直エッジを画面の中心に保つように回転台を利用してカメラを回転させる。オプティカルフローと回転台の角速度などの情報を利用して環境構造を決定することができる。彼らはこの方法で求められた局所地図の注視点(垂直エッジ)の近傍の精度は注視制御を行わない場合に比べて高いと主張している。

Ballard^[3]は Aloimonos の研究を発展させ、能動的視覚と人間の視覚行動との共通点を重要視し、人間の視覚行動をもとにした Animate Vision と名付ける研究枠組を提案した。2台のカメラをマニピュレータの先端に取り付け、人間の眼球運動に似た機能を実現した。

Zheng と Chen^{[40][41]}は能動的視覚の手法を導入し、ビジョンフィードバック、そしてマニプレータ及び被操作物体の状況変化に応じてカメラの視点を能動的に調整することによって、環境モデルを必要としないマニプレータによる物体を操作する方法を提案した。この研究には3つの要素がある。

1. ビジョンフィードバックによる物体の操作
2. 物体の操作に適するカメラの視点の選択
3. 選択された視点へのカメラの移動の制御

彼らは特に能動的視覚のキーである注視点の選択方法及びそ

の視点へカメラを移動させるについて考査した。彼らは積木世界における物体の操作の環境において良い視点を次のように定義した。

1. 物体の3次元の位置関係ははっきり分かる視点。物体を操作する場合、被操作物体とロボットハンドの間の距離ははっきり分かる視点
2. 物体の詳細をはっきりするような視点

物体の操作の場合、視点(1)に移動するために調べる対象となる2つの物体をほぼ画面中心に保ちながらその二つの物体の画像上の距離が増大する方向にカメラ移動させる。物体が画面の端になると、その物体が画面の中心に来るようにカメラを回転させる。そして画面上の物体の大きさが変化するとその物体の大きさが元に戻るように物体とカメラ間の距離を調整する。視点(2)に移動するために物体とカメラ間の距離及びカメラの解像度を調整する。その他、物体の操作及びカメラを移動するとき、オクルジョンを回避する方法などについての研究もなされた。図3はこのハンドアイシステムによる積木の操作の様子、図4はハンドアイシステムの構成を示し、図5は物体間の位置関係が良く分かる視点へカメラを移動させる実験の様子を示す。

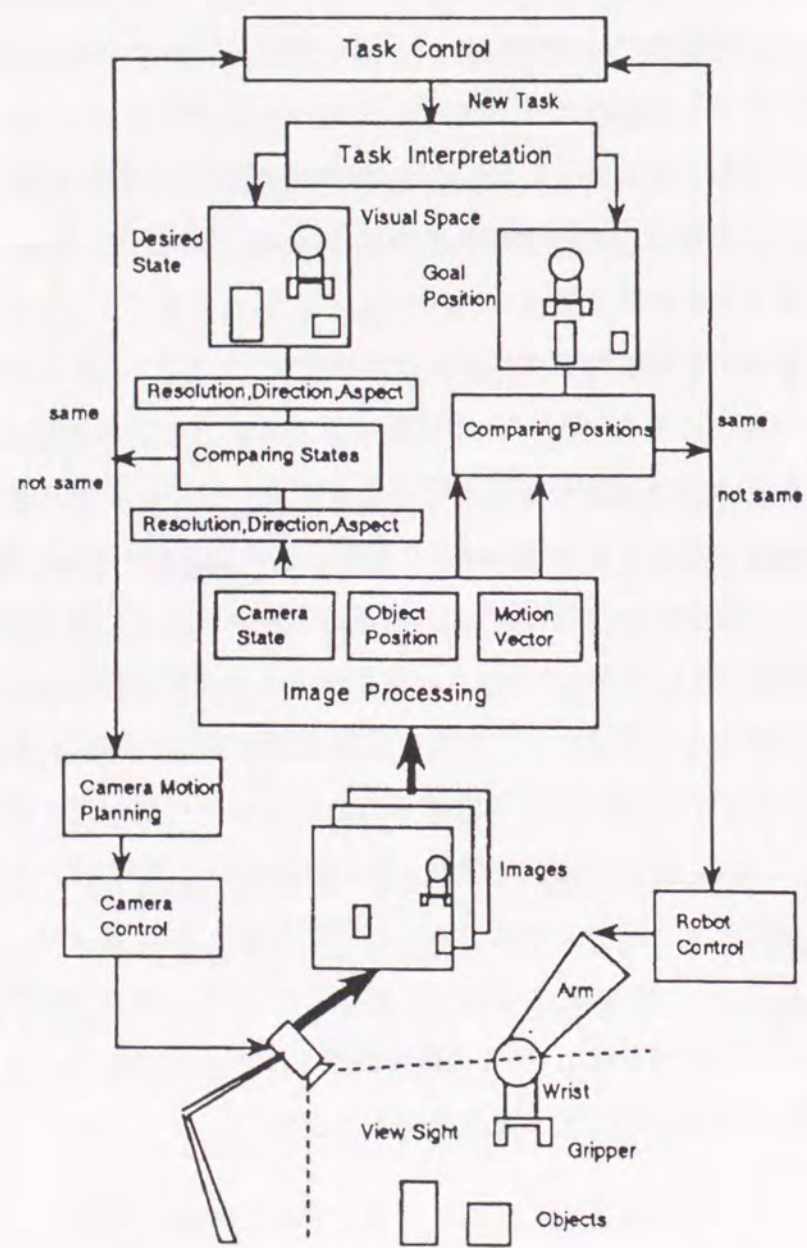


Figure 3 The overview of the hand-eye system.

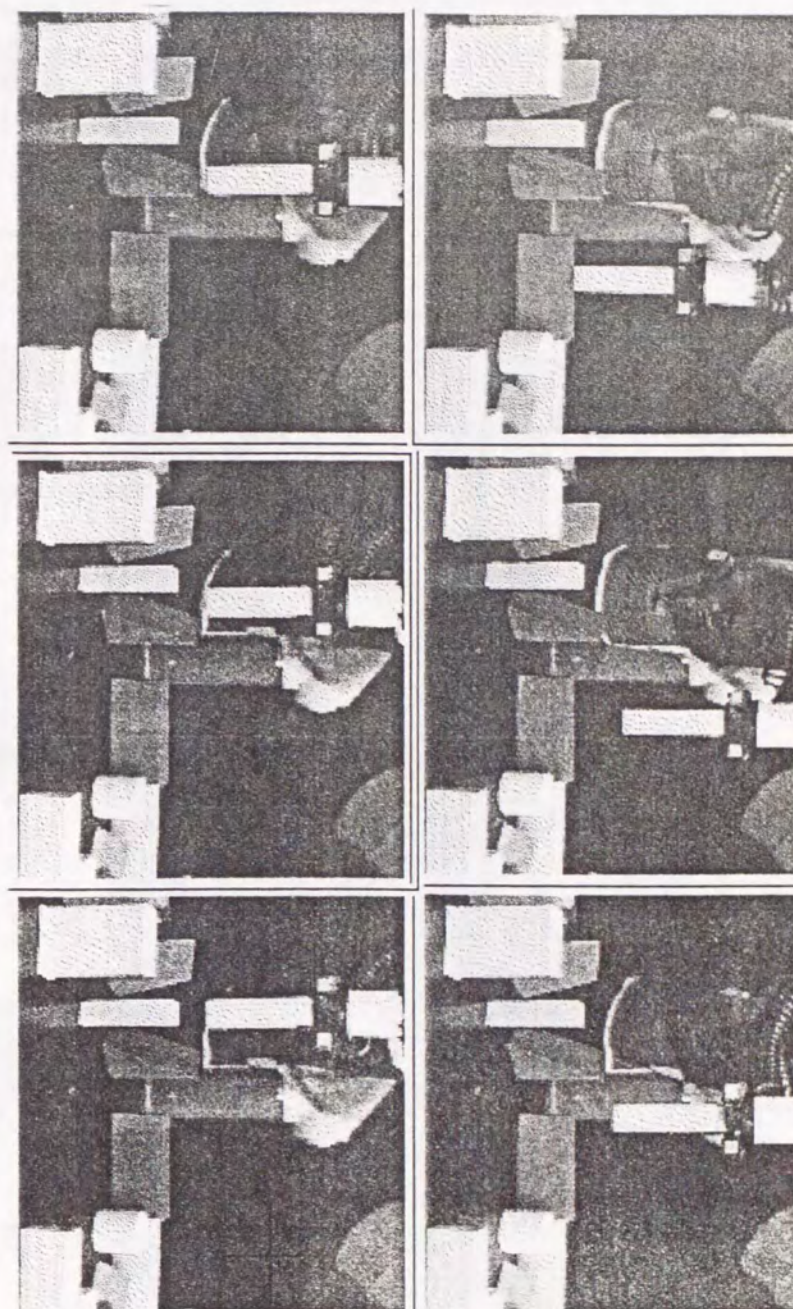


Figure 4 Moving hand to locate the object onto another one.

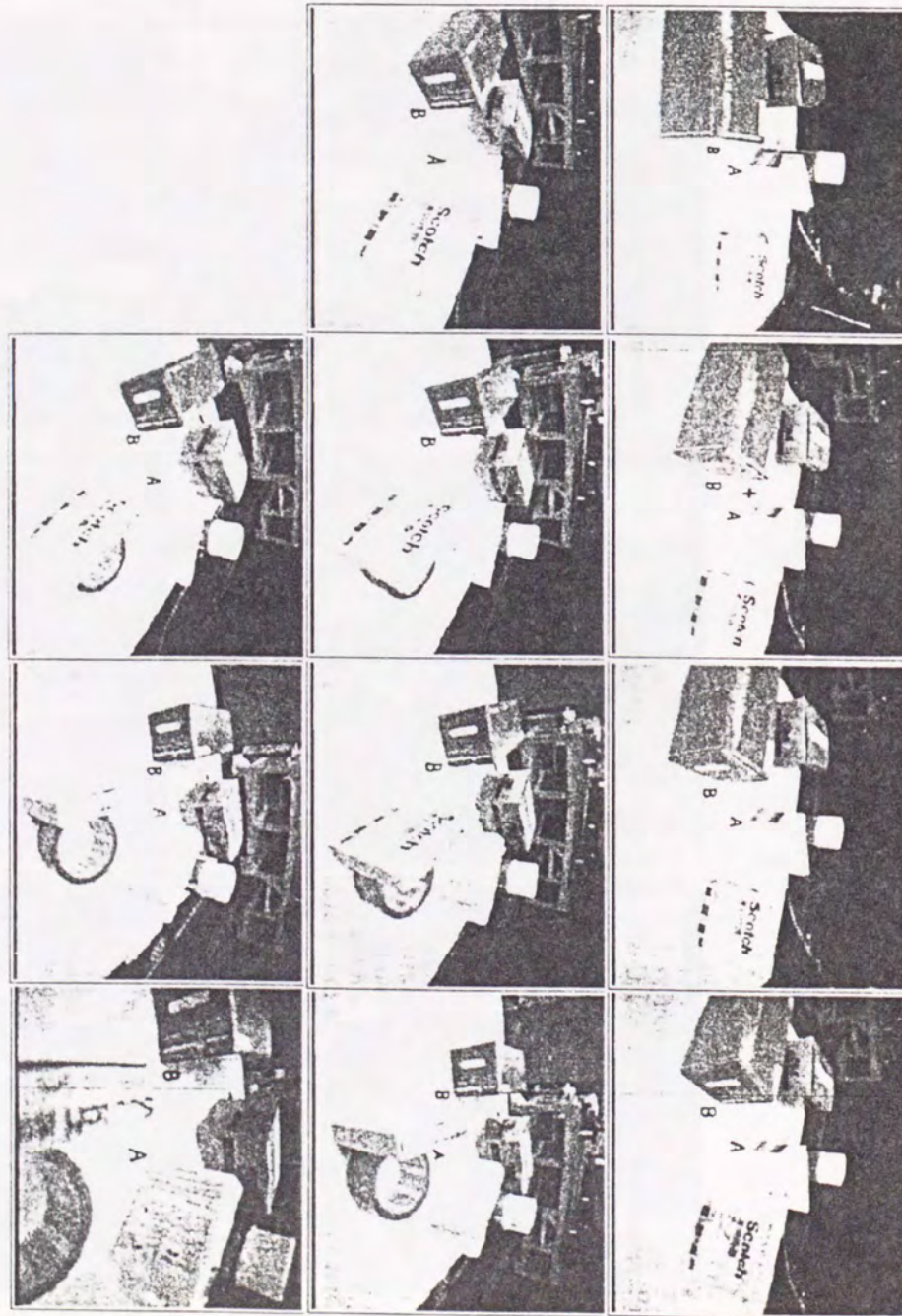


Figure 5 Moving camera to get a good view of objects.

能動的視覚は受動的視覚に比べていくつかの利点があるが能動的視覚の手法を利用するには以下の条件が必要である。

1. カメラの正確な運動の情報獲得できること
2. カメラの運動が自由に制御できること
3. 注視点が適当に選択すること

しかし現実にはマニプレータのようなロボット以外、移動ロボットの場合カメラの正確な運動情報を知ることは困難である。従って視覚情報によってカメラの運動を抽出する必要がある。そして移動ロボットの場合、必ずしもカメラの運動を制御できない。

以上の原因で受動的視覚は能動的視覚によって完全に替えられることはできない。特に移動ロボットの場合、受動的視覚は依然必要である。

3

SVDによる行列の分解 及びノイズの除去

この章では *Singular Value Decomposition* という線形代数のアルゴリズムを利用して、行列の分解とノイズの除去を行う方法とその原理を紹介する。

3.1 Singular Value Decomposition アルゴリズムとその性質

線形代数には *Singular Value Decomposition (SVD)*^[9] というアルゴリズムが存在する。このアルゴリズムによると、 $m \times n$ ($m \geq n$) の行列 A は次のように分解することができる。

$$A = U \Sigma V^T \quad (3.1)$$

ここで U と V はそれぞれ $m \times m$ と $n \times n$ の行列で、 Σ は $n \times n$ の対角行列である。 U, V, Σ は以下の条件を満足する。

$$U^T U = V^T V = V V^T = I_n \quad (3.2)$$

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0 \quad (3.3)$$

U は $A A^T$ の n 個の最も大きい値を持つ固有値に対応している正規直交固有ベクトルにより構成され、 V は行列 $A^T A$ の正規直交固有ベクトルにより構成される。行列 Σ の対角線上の要素は $A^T A$ の固有値の正の平方根であり、*singular value* と呼ぶ。行列 A の階数 (rank) は r であるとき、

$$\sigma_{r+1} = \sigma_{i+2} = \dots = \sigma_n = 0 \quad (3.4)$$

(3.1) の行列の分解は **singular value decomposition** と呼ぶ。
 $m < n$ の $m \times n$ の行列 A の singular value decomposition を求める時、まず次のような A の転置行列の分解を求め、

$$A^T = U\Sigma V^T$$

A の singular value decomposition を次のように求める。

$$A = V\Sigma U^T$$

議論を簡単にするために、一般性を失わないで、SVD を適用する行列の行は列より大きい或は等しいと仮定する。

定理 1

任意の行列のすべての要素の 2 乗の和はその行列の singular value の 2 乗の和に等しい。

証明:

(3.1) により、 $m \times n (m \geq n)$ の行列 A は次のように分解でき、

$$A = U\Sigma V^T$$

行列 A の要素は次のように表現できる。

$$a_{ij} = \sum_{k=1}^n u_{ik} \sigma_k v_{jk} \quad (3.5)$$

従って

$$a_{ij}^2 = \sum_{k_1=1}^n \sum_{k_2=1}^n u_{ik_1} u_{ik_2} \sigma_{k_1} \sigma_{k_2} v_{jk_1} v_{jk_2}$$

行列 A の全ての要素の自乗の和が次の式で計算できる。

$$\begin{aligned} S &= \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \\ &= \sum_{i=1}^m \sum_{j=1}^n \sum_{k_1=1}^n \sum_{k_2=1}^n u_{ik_1} u_{ik_2} \sigma_{k_1} \sigma_{k_2} v_{jk_1} v_{jk_2} \\ &= \sum_{k_1=1}^n \sum_{k_2=1}^n (\sigma_{k_1} \sigma_{k_2} (\sum_{i=1}^m u_{ik_1} u_{ik_2})) (\sum_{j=1}^n v_{jk_1} v_{jk_2}) \end{aligned} \quad (3.6)$$

(3.2) により次の式が成立する。

$$\sum_{k=1}^n u_{ki} u_{kj} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases} \quad (3.7)$$

$$\sum_{k=1}^n v_{ki} v_{kj} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$

(3.6)(3.7) を (3.5) に代入して次の結果が得られる。

$$\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \sum_{k=1}^n \sigma_k^2 \quad (3.9)$$

証明終了。

3.2 SVD による行列の分解

この節では SVD を利用して、階数(rank) は r である $m \times n$ ($m \geq n$) の行列 A を $m \times r$ の行列と $r \times n$ の行列の積に分解する方法を紹介する。

SVDのアルゴリズムを利用してAを次のように分解できる.

$$A = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix}$$

Aの階数はrであるとき, $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_n = 0$ が成立するので, 次の式が得られる.

$$A = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ u_{21} & u_{22} & \dots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mr} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1r} & v_{2r} & \dots & v_{nr} \end{pmatrix}$$

従って, 行列Aは次のように $m \times r$ の行列M'と $r \times n$ の行列S'の積に分解することができる.

$$A = M'S' \quad (3.10)$$

ここで

$$M' = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ u_{21} & u_{22} & \dots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mr} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{pmatrix}$$

$$S' = \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1r} & v_{2r} & \dots & v_{nr} \end{pmatrix}$$

任意の階数がrの $r \times r$ の行列Cに対して, 式

$$M'S' = M'CC^{-1}S' = (M'C)(C^{-1}S') \quad (3.11)$$

が成立するので, 階数がrである $m \times n$ の行列から $m \times r$ の行列と $r \times n$ の行列の積への分解はユニックではない.

3.3 SVDによるノイズの除去

この節ではノイズが存在するとき, 行列を分解するとともにノイズを除去する方法とその原理を紹介する.

ノイズを含んだ行列Aは次のように表現できる.

$$A = \hat{A} + E \quad (3.12)$$

ここで, \hat{A} は行列Aの元の成分で, 階数はrとする. EはAに含んだノイズ成分である.

SVDのアルゴリズムを用いてAと \hat{A} を次のように分解する.

$$A = U\Sigma V^T \quad (3.13)$$

$$\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^T \quad (3.14)$$

従って, A と \hat{A} の要素は次のように表現できる.

$$a_{ij} = \sum_{k=1}^n u_{ik} \sigma_k v_{jk} \quad (3.15)$$

$$\hat{a}_{ij} = \sum_{k=1}^r \hat{u}_{ik} \hat{\sigma}_k \hat{v}_{jk} \quad (3.16)$$

ノイズ行列 E の要素は次のように表現できる.

$$\begin{aligned} e_{ij} &= a_{ij} - \hat{a}_{ij} \\ &= \sum_{k=1}^n u_{ik} \sigma_k v_{jk} - \sum_{k=1}^r \hat{u}_{ik} \hat{\sigma}_k \hat{v}_{jk} \\ &= \sum_{k=1}^r (u_{ik} \sigma_k v_{jk} - \hat{u}_{ik} \hat{\sigma}_k \hat{v}_{jk}) + \sum_{k=r+1}^n u_{ik} \sigma_k v_{jk} \end{aligned} \quad (3.17)$$

ノイズ成分は少ないと仮定する. この場合 e_{ij} は次のように近似できる.

$$\begin{aligned} e_{ij} &= \sum_{k=1}^r (u_{ik} \sigma_k \Delta v_{jk} + u_{ik} \Delta \sigma_k v_{jk} + \Delta u_{ik} \sigma_k v_{jk}) \\ &\quad + \sum_{k=r+1}^n u_{ik} \sigma_k v_{jk} \\ &= \sum_{k=1}^r u_{ik} \sigma_k \Delta v_{jk} + \sum_{k=1}^r u_{ik} \Delta \sigma_k v_{jk} \\ &\quad + \sum_{k=1}^r \Delta u_{ik} \sigma_k v_{jk} + \sum_{k=r+1}^n u_{ik} \sigma_k v_{jk} \end{aligned} \quad (3.18)$$

ここで,

$$\begin{cases} \Delta u_{jk} = u_{jk} - \hat{u}_{jk} \\ \Delta \sigma_k = \sigma_k - \hat{\sigma}_k \\ \Delta v_{jk} = v_{jk} - \hat{v}_{jk} \end{cases}$$

従って, ノイズ行列 E の全ての成分の自乗の和 S_{e^2} は次の式で計算できる.

$$\begin{aligned} S_{e^2} &= \sum_{i=1}^m \sum_{j=1}^n e_{ij}^2 \\ &= \sum_{i=1}^m \sum_{j=1}^n \left(\sum_{k=1}^r u_{ik} \sigma_k \Delta v_{jk} + \sum_{k=1}^r u_{ik} \Delta \sigma_k v_{jk} \right. \\ &\quad \left. + \sum_{k=1}^r \Delta u_{ik} \sigma_k v_{jk} + \sum_{k=r+1}^n u_{ik} \sigma_k v_{jk} \right)^2 \end{aligned} \quad (3.19)$$

(3.19) を展開し, 整理すると, 次の式が得られる.

$$\begin{aligned} S_{e^2} &= \sum_{k_1=1}^r \Delta \sigma_{k_1} \left(\sum_{k_2=1}^r \Delta \sigma_{k_2} \left(\sum_{i=1}^m u_{ik_1} u_{ik_2} \left(\sum_{j=1}^n v_{jk_1} v_{jk_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \sigma_{k_1} \left(\sum_{k_2=1}^r \sigma_{k_2} \left(\sum_{i=1}^m \Delta u_{ik_1} u_{ik_2} \left(\sum_{j=1}^n v_{jk_1} v_{jk_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \sigma_{k_1} \left(\sum_{k_2=1}^r \sigma_{k_2} \left(\sum_{j=1}^n \Delta v_{jk_1} v_{jk_2} \left(\sum_{i=1}^m u_{ik_1} u_{ik_2} \right) \right) \right) \\ &\quad + \sum_{k_1=r+1}^n \sigma_{k_1} \left(\sum_{k_2=r+1}^n \sigma_{k_2} \left(\sum_{i=1}^m u_{ik_1} u_{ik_2} \left(\sum_{j=1}^n v_{jk_1} v_{jk_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \Delta \sigma_{k_1} \left(\sum_{k_2=1}^r \sigma_{k_2} \left(\sum_{j=1}^n v_{jk_1} \Delta v_{jk_2} \left(\sum_{i=1}^m u_{ik_1} u_{ik_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \Delta \sigma_{k_1} \left(\sum_{k_2=1}^r \sigma_{k_2} \left(\sum_{i=1}^m u_{ik_1} \Delta u_{ik_2} \left(\sum_{j=1}^n v_{jk_1} v_{jk_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \Delta \sigma_{k_1} \left(\sum_{k_2=r+1}^n \sigma_{k_2} \left(\sum_{i=1}^m u_{ik_1} \Delta u_{ik_2} \left(\sum_{j=1}^n v_{jk_1} v_{jk_2} \right) \right) \right) \\ &\quad + \sum_{k_1=1}^r \sigma_{k_1} \left(\sum_{k_2=1}^r \sigma_{k_2} \left(\sum_{i=1}^m u_{ik_1} \Delta u_{ik_2} \left(\sum_{j=1}^n \Delta v_{jk_1} v_{jk_2} \right) \right) \right) \end{aligned}$$

$$\begin{aligned}
& + \sum_{k1=1}^r \sigma_{k1} \left(\sum_{k2=r+1}^n \sigma_{k2} \left(\sum_{j=1}^m \Delta v_{jk1} v_{jk2} \left(\sum_{i=1}^m u_{ik1} u_{ik2} \right) \right) \right) \\
& + \sum_{k1=1}^r \sigma_{k1} \left(\sum_{k2=r+1}^n \sigma_{k2} \left(\sum_{i=1}^m \Delta u_{ik1} u_{ik2} \left(\sum_{j=1}^n v_{jk1} v_{jk2} \right) \right) \right)
\end{aligned} \quad (3.20)$$

(3.2)により次の式が成立する.

$$\sum_{k=1}^n u_{ki} u_{kj} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases} \quad (3.21)$$

$$\sum_{k=1}^n v_{ki} v_{kj} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{otherwise.} \end{cases} \quad (3.22)$$

従って(3.20)から次の式が得られる.

$$\begin{aligned}
S_{e^2} &= \sum_{k=1}^r \Delta \sigma_k^2 + \sum_{k=1}^r \sigma_k^2 \left(\sum_{i=1}^m \Delta u_{ik} u_{ik} \right) + \sum_{k=1}^r \sigma_k^2 \left(\sum_{j=1}^n \Delta v_{jk} v_{jk} \right) \\
&+ \sum_{k=r+1}^n \sigma_k^2 + \sum_{k=1}^r \Delta \sigma_k \sigma_k \left(\sum_{j=1}^n v_{jk} \Delta v_{jk} \right) \\
&+ \sum_{k=1}^r \Delta \sigma_k \sigma_k \left(\sum_{i=1}^m u_{ik} \Delta u_{ik} \right) \\
&+ \sum_{k1=1}^r \sigma_{k1} \left(\sum_{k2=1}^r \sigma_{k2} \left(\sum_{i=1}^m u_{ik1} \Delta u_{ik2} \left(\sum_{j=1}^n \Delta v_{jk1} v_{jk2} \right) \right) \right)
\end{aligned} \quad (3.23)$$

この式の中のパラメータのうち行列 \hat{A} の要素と関連のないものは $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_n$ である. これらのパラメータを0に設定すると, S_{e^2} の値を小さくなる. 従って, $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_n$ を0に設定することにより, ノイズを部分的に除去することができる. 行列 A の最

良近似は次のように求めることができる.

$$A = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mn} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_n \end{pmatrix} \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix}$$

3.2に紹介してある方法と同じように行列を次のように $m \times r$ の行列 M' と $r \times n$ の行列 S' の積に分解することができる.

$$A = M'S' \quad (3.24)$$

ここで

$$M' = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ u_{21} & u_{22} & \dots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mr} \end{pmatrix} \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{pmatrix}$$

$$S' = \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1r} & v_{2r} & \dots & v_{nr} \end{pmatrix}$$

この操作の直観的な意味はこうである. $\sigma_1, \sigma_2, \dots, \sigma_n$ は $A^T A$ の固有値の平方根である. ノイズは存在しないとき, $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_n$ は0である. 従って, $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_n$ による生成される成分は全てノイズの影響によるもので, $\sigma_{r+1}, \sigma_{r+2}, \dots, \sigma_n$ を0に設定することにより, ノイズを除去することができる.

従って行列の階数を知れば, *singular value decomposition* のアルゴリズムを利用して行列を分解し, そして行列の階数個以降の *singular value* を0に設定することにより, ノイズを除去することができ, 行列を分解することができる.

4

連続画像を利用する 動物体の抽出

従来の structure from motion の研究のほとんどは静止環境或いは剛体運動を前提とする。しかし、実際の環境に動物体は存在するし、シーンの中に別々の方向で異なる速度で移動する物体は存在するが多い。このような環境において、静止環境或は剛体運動を前提とする手法を利用して環境構造を推定するために、シーンの中の静止物体と動物体を分類する必要がある。

この章では連続の画像列から *singular value decomposition* のアルゴリズムを利用して静止物体と動物体を分離する方法を紹介する。

我々は環境に静止物体が多数存在し、動物体の数は静止物体に比べて非常に少ないと仮定し、カメラは並進運動と回転運動をしながら連続の画像を撮り、シーンからカメラへの投影は平行射影であると仮定する。そして、連続画像の各フレームにおいて物体の特徴点は既に抽出しており、特徴点のフレーム間の対応付けは既に求められていると仮定する。

議論を簡単にするために、我々は環境内の物体運動とカメラの運動は水平面に限定し、全ての画像特徴は垂直エッジで、occlusion は存在しないと仮定する。この場合、我々は画像の中の一本の水平走査線だけを考える。

環境の世界座標系とカメラ座標系は図6に示されるように設定する。

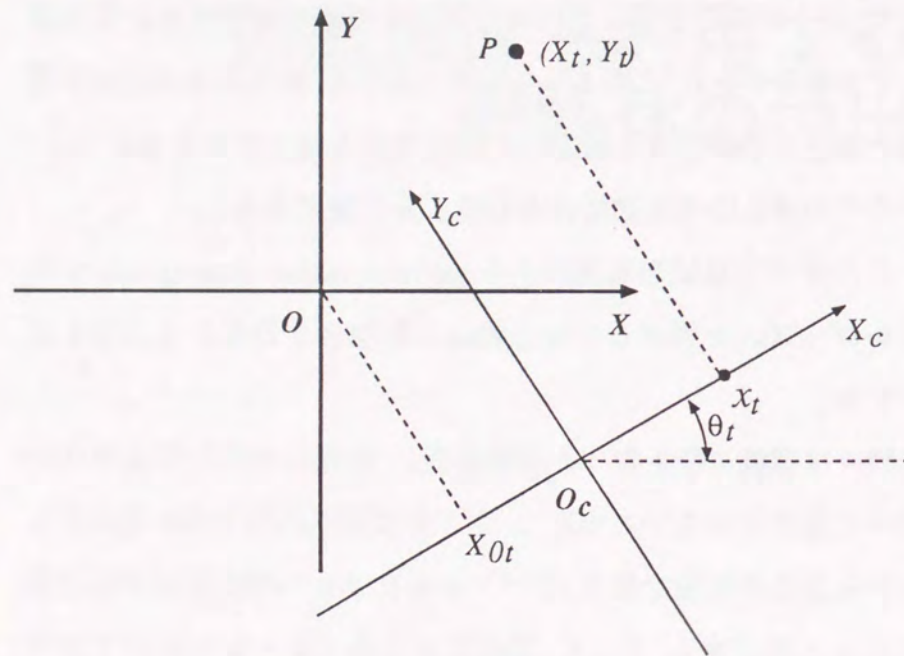


Figure 6 The world coordinate system and camera coordinate system.

ある物体の特徴点Pのt時刻で撮像されたの画像での上の座標は次の式で表される。

$$x_t = X_t \cos \theta_t + Y_t \sin \theta_t + X_{0t} \quad (4.1)$$

ここで (X_t, Y_t) は特徴点Pのt時刻で世界座標系での座標で、 θ_t はt時刻でのカメラの光軸が世界座標系のY軸に対する回転角度で、 X_{0t} は世界座標系の原点のカメラ座標系でのX座標である。

(X_{t_0}, Y_{t_0}) はPの t_0 時刻での位置とする。動物体の動きによって生じるカメラ画面上の動物体の像と静止物体の像の動きの間のずれ e_t は次の式で計算できる。

$$\begin{aligned} e_t &= X_t \cos \theta_t + Y_t \sin \theta_t + X_{0t} - (X_{t_0} \cos \theta_t + Y_{t_0} \sin \theta_t + X_{0t}) \\ &= (X_t - X_{t_0}) \cos \theta_t + (Y_t - Y_{t_0}) \sin \theta_t \end{aligned} \quad (4.2)$$

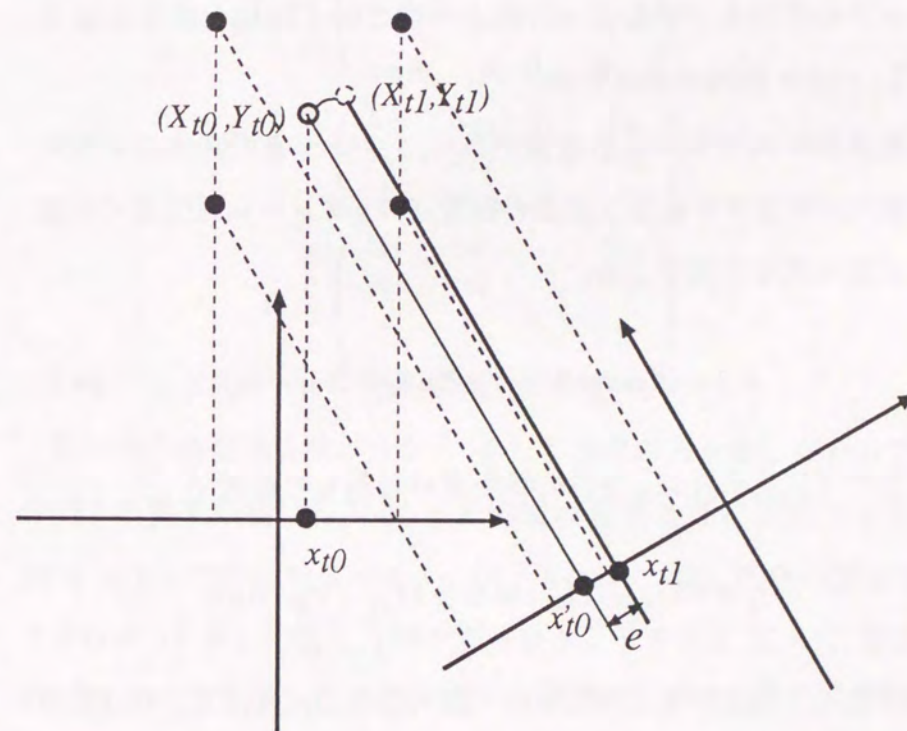


Figure 7 The difference of the position of the projection caused by the movement of the feature point.

静止物体の場合,

$$\begin{cases} X_t = X_{t_0} \\ Y_t = Y_{t_0} \end{cases}$$

従って, 全てのフレームにおいて, e_t の値は 0 になる. 物体が移動する場合, 以下のような特殊な運動パターンを除き,

$$(X_t - X_{t_0}) \cos \theta_t + (Y_t - Y_{t_0}) \sin \theta_t = 0$$

カメラの画面上で動物体の特徴点の像は静止物体と異なる動きをし, e_t は 0 以外の値になる.

画像列に m フレームの画像があり, 1 フレームの画像に n 個の特徴点が存在する場合, j 番目の特徴点の i フレーム画像上での座標は次の式で表現できる.

$$x_{ij} = X_{j0} \cos \theta_i + Y_{j0} \sin \theta_i + X_{0i} + e_{ij} \quad (4.3)$$

ここで, (X_{j0}, Y_{j0}) は j 番目の特徴点の時刻 t_0 の位置で,

$$e_{ij} = (X_{ji} - X_{j0}) \cos \theta_i + (Y_{ji} - Y_{j0}) \sin \theta_i$$

は動物体の動きが生じた特徴点の像の動きで, (X_{ji}, Y_{ji}) は j 番目の特徴点の時刻 t_i の位置である.

画像列を次のように表現できる.

$$\mathbf{I} = \mathbf{MS} + \mathbf{E} \quad (4.4)$$

ここで \mathbf{M} は t_0 時刻の物体の特徴点の位置で, \mathbf{M} はカメラの運動パラメータで, \mathbf{E} は物体の動きによって生じる特徴点の像と物体が静止する場合の特徴点の像の間のずれである.

$$\mathbf{I} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$$

$$\mathbf{M} = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 & T_1 \\ \cos \theta_2 & \sin \theta_2 & T_2 \\ \vdots & \vdots & \vdots \\ \cos \theta_m & \sin \theta_m & T_m \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} X_{01} & X_{02} & \dots & X_{0n} \\ Y_{01} & Y_{02} & \dots & Y_{0n} \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

$$\mathbf{E} = \begin{pmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & e_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & \dots & e_{mn} \end{pmatrix}$$

動物体の特徴点は数が少なく, そして急激の動きをしないので, 行列 \mathbf{I} の要素の値は行列 \mathbf{M} と \mathbf{S} の積の要素とほぼ同じである. \mathbf{M} は $m \times 3$ の行列で, \mathbf{S} は $3 \times n$ の行列なので, 線形代数の理論により行列 \mathbf{M} と \mathbf{S} の積の行列の階数は 3 以下である. 従って, 物体の特徴点の特殊な分布 (例えば全ての特徴点が一直線上にある), 或いはカメラが特殊な運動 (例えば並進運動のみ) をする場合を除き, 行列 \mathbf{MS} の階数は 3 である.

$$\text{rank}(\mathbf{MS}) = 3 \quad (4.5)$$

この条件を利用して、 E をノイズ行列と見なし、3章で紹介した手法を利用してノイズを除去し MS の最良の近似 \hat{MS} を求めることができる。従って、次の式を利用して E を計算することができる。

$$E \approx I - \hat{MS} \quad (4.6)$$

E の中で動物体の特徴点に対応する要素は0以外の値になり、静止物体の特徴点に対応するものは0になるはずである。実際に画像ノイズの影響によって静止物体の特徴点に対応する要素でも非0値を持つ場合がある。しかし、その値の絶対値は動物体の特徴点に対応する要素に比べてずっと小さいはずである。

E の j 列目の要素は j 番目の物体の特徴点に対応するので、 E の各列の要素の自乗の和を計算する。

$$Se = (Se_1 \quad Se_2 \quad \dots \quad Se_n) \quad (4.7)$$

ここで、 $Se_j = \sum_{i=1}^m e_{ij}^2$ 。

Se のヒストグラムを作り、それを利用して Se の要素を2種類に分類し、大きい値を持つ Se の要素に対応する特徴点は動物体の特徴点で、それ以外の特徴点は静止物体の特徴点である。

5

深さ範囲が狭いシーンの構造の推定

この章では物体の大きさがカメラとの距離に比べて非常に小さい、つまり深さ範囲が狭い場合において、連続画像からシーンの構造とカメラ運動を推定する方法を紹介する。シーンの中に動物体は存在しなく、そして画像列の各フレームにおいて特徴点はすでに抽出しており、特徴点のフレーム間の対応付けは既に求められていると仮定する。

議論を簡単にするために、我々は環境内の物体運動とカメラの運動は水平面に限定し、全ての画像特徴は垂直エッジで、occlusionは存在しないと仮定する。この場合、我々は画像の中の一本の水平走査線だけを考える。

カメラの移動距離が短く、且つ回転角度が小さい場合、物体の特徴点の深さ (depth) はほとんど変化しないので、シーンからカメラへの投影は平行射影で近似できる。この場合、平行射影の線形性を利用し、*singular value decomposition* のアルゴリズムによるノイズが含まれた行列を分解する方法を用いてシーンの構造を推定できる。しかし、カメラの移動が大きいとき、平行射影のパラメータの変化が大きいため、カメラへの投影は一つの平行射影で近似できなくなる。この場合、画像列を分割することにより、分割された各サブ画像列が撮像される過程におけるカメラの運動は小さくなり、シーンからカメラへの投影は平行射影で近似してシーンの構造を求める。各サブ画像列から求められた物体の形状は不変という条件を利用してサブ画像列間の平行射影のパラメータの

変化を求め、物体の形状とカメラ運動を推定する。

5.1 行列の分解による形状情報と運動情報の分離

図8が示すように、世界座標系 O - XY の原点 O を物体の重心に選び、カメラ座標系 O_c - X_cY_c の原点 O_c をレンズの焦点に固定し、レンズの光軸を Y 軸とする。

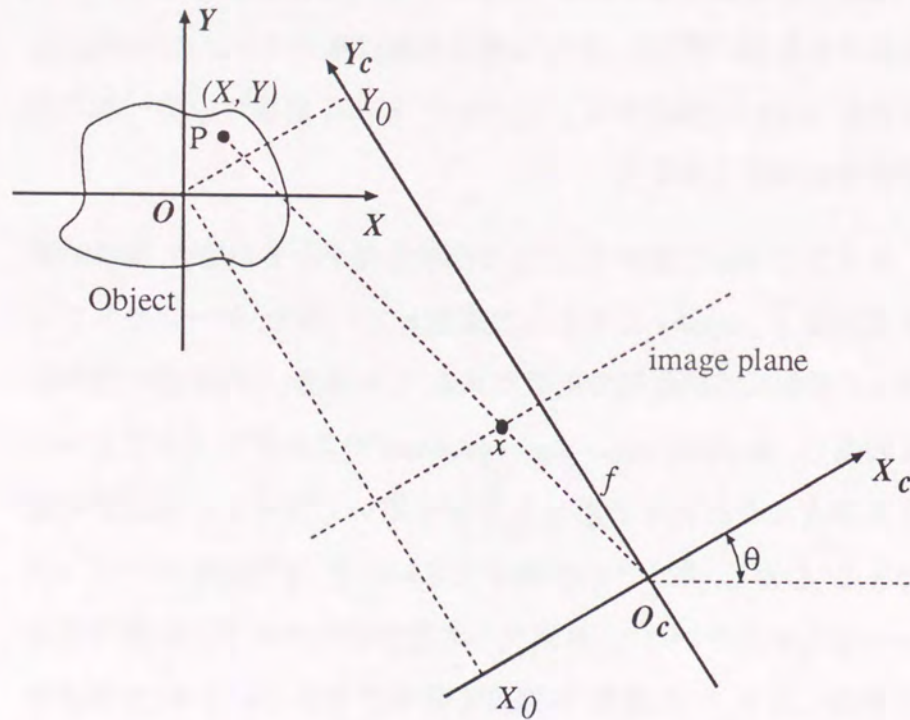


Figure 8 The camera coordinate system.

カメラ座標系の X_c 軸と世界座標系の X 軸間の角度は θ とし、世界座標系の原点 O がカメラ座標系での座標は (X_0, Y_0) とする。物

体の特徴点 P がカメラ座標系での座標は次のように表現できる。

$$\begin{cases} X_c = X \cos \theta + Y \sin \theta + X_0 \\ Y_c = -X \sin \theta + Y \cos \theta + Y_0 \end{cases} \quad (5.1)$$

ピンホールカメラモデルにより、 P のカメラの画面上の像の位置は次のように計算できる。

$$x = f \frac{X_c}{Y_c} = f \frac{X \cos \theta + Y \sin \theta + X_0}{-X \sin \theta + Y \cos \theta + Y_0} \quad (5.2)$$

もし物体の大きさが物体とカメラ間の距離に比べて十分小さければ、つまり次の式が成立する場合、

$$\sqrt{X^2 + Y^2} \ll Y_0 \quad (5.3)$$

(5.2) を次のように書き換え、

$$x = f \frac{X \cos \theta + Y \sin \theta + X_0}{Y_0} (1 + \Delta Y) \quad (5.4)$$

ここで

$$\begin{aligned} |\Delta Y| &= \frac{|X \sin \theta - Y \cos \theta|}{|-X \sin \theta + Y \cos \theta + Y_0|} \\ &< \frac{\sqrt{X^2 + Y^2}}{Y_0 - \sqrt{X^2 + Y^2}} \\ &= \frac{\frac{\sqrt{X^2 + Y^2}}{Y_0}}{1 - \frac{\sqrt{X^2 + Y^2}}{Y_0}} \ll 1.0 \end{aligned}$$

ΔY を省略して、次の式が得られる。

$$x = f \frac{X \cos \theta + Y \sin \theta + X_0}{Y_0} \quad (5.5)$$

$\frac{\sqrt{X^2+Y^2}}{Y_0}$ は物体の大きさと物体とカメラ間の平均距離の比率を表す。我々はこの比率を深さ範囲と定義する。この比率が大きい場合、深さ範囲は広いといい、逆に、この比率が小さいと深さ範囲は狭いという。

物体の特徴点の数を n とし、世界座標系の原点を物体の重心に選んだので、次の式が得られる。

$$\begin{cases} \sum_{i=1}^n X_i = 0 \\ \sum_{i=1}^n Y_i = 0 \end{cases} \quad (5.6)$$

画像内の全ての特徴点の座標の和を計算して、

$$\sum_{i=1}^n x = f \frac{\cos \theta \sum_{i=1}^n X_i + \sin \theta \sum_{i=1}^n Y_i + nX_0}{Y_0} \quad (5.7)$$

(5.6) を (5.7) に代入して、次の式が得れる。

$$\frac{X_0}{Y_0} = \frac{\sum_{i=1}^n x_i}{nf} \quad (5.8)$$

(5.8) と (5.5) から次の結果が得られる。

$$\frac{X \cos \theta + Y \sin \theta}{Y_0} = x' \quad (5.9)$$

ここで、

$$x' = \frac{x}{f} - \frac{X_0}{Y_0} = \frac{x}{f} - \frac{\sum_{i=1}^n x_i}{nf}$$

(5.9) の右辺の値は既知である。

画像列に m フレームの画像があるとし、(5.9) より次の式が得られる。

$$\mathbf{I}' = \mathbf{M}\mathbf{S} \quad (5.10)$$

ここで、

$$\mathbf{I}' = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$$

$$\mathbf{M} = \begin{pmatrix} \frac{\cos \theta_1}{Y_{01}} & \frac{\sin \theta_1}{Y_{01}} \\ \frac{\cos \theta_2}{Y_{02}} & \frac{\sin \theta_2}{Y_{02}} \\ \vdots & \vdots \\ \frac{\cos \theta_m}{Y_{0m}} & \frac{\sin \theta_m}{Y_{0m}} \end{pmatrix}$$

$$\mathbf{S} = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \end{pmatrix}$$

式の中、 θ_i, Y_{0i} ($i = 1, 2, \dots, m$) は i 番目の画像が撮像される時のカメラの回転角度とカメラ座標系での世界座標系の原点の Y 座標であり、 (X_j, Y_j) ($j = 1, 2, \dots, n$) は j 番目の特徴点の世界座標系での位置である。

実画像にはノイズが存在するので、(5.10) は次のようになる。

$$\mathbf{I}' = \mathbf{M}\mathbf{S} + \mathbf{E} \quad (5.11)$$

ここで、 \mathbf{E} はノイズ行列である。

\mathbf{M} は $m \times 2$ の行列で、 \mathbf{S} は $2 \times n$ の行列なので、線形代数の理論により行列 \mathbf{M} と \mathbf{S} の行列の積の行列の階数は 2 以下である。物

体の特徴点の特殊な分布 (例えば全ての特徴点が一直線上にある), 或いはカメラが特殊な運動 (例えば並進運動のみ) をする場合を除いて, 行列 $\mathbf{M}\mathbf{S}$ の階数は2である. 3章で紹介した方法を利用してノイズを除去し, \mathbf{I}' を $m \times 2$ の行列と $2 \times n$ の行列の積に分解することができる.

$$\mathbf{I}' = \hat{\mathbf{M}}\hat{\mathbf{S}} \quad (5.12)$$

ここで

$$\hat{\mathbf{M}} = \begin{pmatrix} \hat{m}_{11} & \hat{m}_{12} \\ \hat{m}_{21} & \hat{m}_{22} \\ \vdots & \vdots \\ \hat{m}_{m1} & \hat{m}_{m2} \end{pmatrix}$$

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{s}_{11} & \hat{s}_{21} & \dots & \hat{s}_{n1} \\ \hat{s}_{12} & \hat{s}_{22} & \dots & \hat{s}_{n2} \end{pmatrix}$$

任意の階数が2である 2×2 の行列 \mathbf{C} に対して, 次の式が成立する.

$$\hat{\mathbf{M}}\hat{\mathbf{S}} = \hat{\mathbf{M}}\mathbf{C}\mathbf{C}^{-1}\hat{\mathbf{S}} = (\hat{\mathbf{M}}\mathbf{C})(\mathbf{C}^{-1}\hat{\mathbf{S}})$$

従って, ある \mathbf{C} が存在し,

$$\begin{cases} \mathbf{M} = \hat{\mathbf{M}}\mathbf{C} \\ \mathbf{S} = \mathbf{C}^{-1}\hat{\mathbf{S}} \end{cases} \quad (5.13)$$

\mathbf{C} と \mathbf{C}^{-1} を次のように表現し,

$$\begin{cases} \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \\ \mathbf{C}^{-1} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \end{cases} \quad (5.14)$$

次の式が得れる.

$$\begin{cases} \frac{\cos \theta_i}{Y_{0i}} = c_{11}\hat{m}_{i1} + c_{21}\hat{m}_{i2} \\ \frac{\sin \theta_i}{Y_{0i}} = c_{12}\hat{m}_{i1} + c_{22}\hat{m}_{i2} \end{cases} \quad i = 1, 2, \dots, m \quad (5.15)$$

$$\begin{cases} X_j = d_{11}\hat{s}_{1j} + d_{12}\hat{s}_{2j} \\ Y_j = d_{21}\hat{s}_{1j} + d_{22}\hat{s}_{2j} \end{cases} \quad j = 1, 2, \dots, n \quad (5.16)$$

$\mathbf{C}\mathbf{C}^{-1} = \mathbf{I}_2$ なので, 次の方程式が得られる.

$$\begin{cases} c_{11}d_{11} + c_{12}d_{21} = 1 \\ c_{11}d_{12} + c_{12}d_{22} = 0 \\ c_{21}d_{11} + c_{22}d_{21} = 0 \\ c_{21}d_{12} + c_{22}d_{22} = 1 \end{cases} \quad (5.17)$$

物体の形状情報 (X_j, Y_j) とカメラの運動情報 (θ_i, X_{0i}, Y_{0i}) を求めるには, 変数 $c_{11}, c_{12}, c_{21}, c_{22}$ 及び $d_{11}, d_{12}, d_{21}, d_{22}$ を決定する必要がある. しかし, 方程式 (5.17) の数は未知数の数より少ないので, 物体の形状とカメラ運動を推定するには, 他の制約条件が必要である.

5.2 カメラの移動量が少ない場合での物体形状とカメラ運動の決定

物体の大きさは物体とカメラ間の距離に比べて十分小さく, しかも連続画像を撮像する間, カメラの移動量が少ない場合, 即ちカメラの距離が短く, 回転角度も小さい場合, 物体の全ての特徴

点の深さはほぼ同じで、しかも連続画像を撮像する間ほとんど不変である。連続画像から物体形状とカメラ運動を求めるために、まず5.1節で紹介した方法を用いて連続画像の中に含まれる物体の形状情報とカメラの運動情報を分離する。異なる物体の特徴点の深さの違い及び画像が撮像される間における変化をを省略して、(5.15)を次のように書き換えることができる。

$$\begin{cases} \cos \theta_i = c'_{11} \hat{m}_{i1} + c'_{21} \hat{m}_{i2} \\ \sin \theta_i = c'_{12} \hat{m}_{i1} + c'_{22} \hat{m}_{i2} \end{cases} \quad i = 1, 2, \dots, m \quad (5.18)$$

ここで、

$$\begin{cases} c'_{11} = c_{11} Y_0 \\ c'_{12} = c_{12} Y_0 \\ c'_{21} = c_{21} Y_0 \\ c'_{22} = c_{22} Y_0 \end{cases} \quad (5.19)$$

$\cos^2 \theta_i + \sin^2 \theta_i = 1$ が成立するので、次の制約条件が得られる。

$$\sum_{i=0}^m (\cos^2 \theta_i + \sin^2 \theta_i - 1)^2 \rightarrow \min \quad (5.20)$$

故に、

$$\begin{aligned} & \sum_{i=0}^m (\hat{m}_{i1}^2 (c'_{11}{}^2 + c'_{12}{}^2) + \hat{m}_{i2}^2 (c'_{21}{}^2 + c'_{22}{}^2) \\ & + 2\hat{m}_{i1} \hat{m}_{i2} (c'_{11} c'_{21} + c'_{12} c'_{22}) - 1)^2 \rightarrow \min \end{aligned} \quad (5.21)$$

我々は画像列の最初のフレームの画像が撮像される時のカメラの回転角度 θ_0 を0に設定する。すると、次の式が得られる。

$$\sin \theta_1 = c'_{12} \hat{m}'_{11} + c'_{22} \hat{m}'_{12} = 0$$

故に、

$$c'_{22} = -\frac{\hat{m}_{11}}{\hat{m}_{12}} c'_{12} \quad (5.22)$$

(5.22)を(5.21)に代入して、整理すると次の式が得られる。

$$\sum_{i=1}^m (A_i c'_{11}{}^2 + B_i c'_{12}{}^2 + C_i c'_{21}{}^2 + 2D_i c'_{11} c'_{21} - 1)^2 \rightarrow \min \quad (5.23)$$

ここで、

$$\begin{cases} A_i = \hat{m}_{i1}^2 \\ B_i = \hat{m}_{i1}^2 + \frac{\hat{m}_{11}^2}{\hat{m}_{12}^2} \hat{m}_{i2}^2 - 2\frac{\hat{m}_{11}}{\hat{m}_{12}} \hat{m}_{i1} \hat{m}_{i2} \\ C_i = \hat{m}_{i2}^2 \\ D_i = \hat{m}_{i1} \hat{m}_{i2} \end{cases}$$

(5.23)は非線形の最小自乗問題であり、最小自乗法により次の方程式が得られる。

$$\begin{cases} \sum_{i=1}^m (A_i c'_{11} + D_i c'_{21})(A_i c'_{11}{}^2 + B_i c'_{12}{}^2 + C_i c'_{21}{}^2 + 2D_i c'_{11} c'_{21} - 1) = 0 \\ \sum_{i=1}^m B_i c'_{21} (A_i c'_{11}{}^2 + B_i c'_{12}{}^2 + C_i c'_{21}{}^2 + 2D_i c'_{11} c'_{21} - 1) = 0 \\ \sum_{i=1}^m (C_i c'_{21} + D_i c'_{12})(A_i c'_{11}{}^2 + B_i c'_{12}{}^2 + C_i c'_{21}{}^2 + 2D_i c'_{11} c'_{21} - 1) = 0 \end{cases} \quad (5.24)$$

我々はニュートン法を利用して(5.24)の非線形の方程式の解を求める。ニュートン法を利用するのに必要の解の初期値を得るために、(5.23)の最小自乗問題の中の未知数 c'_{11} , c'_{12} , c'_{21} , $c'_{11} c'_{21}$ を一時的に独立の変数と見なすことにより、次の式が得られる。

$$\sum_{i=1}^m (A_i Z_1 + B_i Z_2 + C_i Z_3 + D_i Z_4 - 1)^2 \rightarrow \min \quad (5.25)$$

ここで

$$\begin{cases} Z_1 = c_{11}'^2 \\ Z_2 = c_{12}'^2 \\ Z_3 = c_{21}'^2 \\ Z_4 = c_{11}'c_{21}' \end{cases} \quad (5.26)$$

我々は最小自乗法を利用して(5.25)で示す問題を解き, Z_1, Z_2, Z_3, Z_4 の値を求める。(5.26)により,

$$c_{12}' = \sqrt{Z_2} \quad (5.27)$$

そして, 次のように行列 C を構築する.

$$C = \begin{pmatrix} Z_1 & Z_4 \\ Z_4 & Z_3 \end{pmatrix}$$

(5.26)により次の式が成立するはずである.

$$C = \begin{pmatrix} c_{11}'^2 & c_{11}'c_{21}' \\ c_{11}'c_{21}' & c_{21}'^2 \end{pmatrix} = \begin{pmatrix} c_{11}' \\ c_{21}' \end{pmatrix} (c_{11}' \quad c_{21}') \quad (5.28)$$

従って, C の階数は1である. Z_1, Z_2, Z_3, Z_4 を計算するとき, これらの変数を独立と仮定しているので, 求められた結果には誤差が存在すると思われる. 従って, 実際に(5.28)は次のようになる.

$$C = \begin{pmatrix} c_{11}' \\ c_{21}' \end{pmatrix} (c_{11}' \quad c_{21}') + E \quad (5.29)$$

ここで, E は誤差行列である.

3章で紹介した方法を利用して C の中に含まれるノイズ成分(誤差)を除去し, 行列 C の分解を行ない, 次の結果が得られる.

$$\begin{pmatrix} c_{11}' \\ c_{21}' \end{pmatrix} (c_{11}' \quad c_{21}') = \hat{C} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} (s_1 \quad s_2) \quad (5.30)$$

従って, ある k が存在し,

$$\begin{cases} c_{11}' = km_1 \\ c_{11}' = \frac{s_1}{k} \\ c_{21}' = km_2 \\ c_{21}' = \frac{s_2}{k} \end{cases} \quad (5.31)$$

(5.31) から次の方程式が得られ,

$$\begin{cases} k^2 = \frac{s_1}{m_1} \\ k^2 = \frac{s_2}{m_2} \end{cases} \quad (5.32)$$

この方程式の最適の解は

$$k = \sqrt{\frac{\frac{s_1}{m_1} + \frac{s_2}{m_2}}{2}} \quad (5.33)$$

であり, この結果を(5.31)に代入することにより, c_{11}', c_{21}' が求められる. c_{11}', c_{21}' と(5.27)で求められた c_{12}' を(5.24)の解の初期値とし, ニュートン法を利用して(5.24)を解き, $c_{11}', c_{12}', c_{21}'$ を求め, そして, (5.22)を利用して c_{22}' を求める.

物体の形状を推定するために, (5.13) 中の C^{-1} を計算する必要がある. (5.17) により,

$$\begin{cases} d_{11} = \frac{c_{22}}{\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}} = \frac{c'_{22}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} Y_0 \\ d_{21} = \frac{-c_{21}}{\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}} = \frac{-c'_{21}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} Y_0 \\ d_{12} = \frac{-c_{12}}{\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}} = \frac{-c'_{12}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} Y_0 \\ d_{22} = \frac{c_{11}}{\begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}} = \frac{c'_{11}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} Y_0 \end{cases} \quad (5.34)$$

(5.18) と (5.8) を利用してカメラの回転運動と並進運動のパラメータが求められる.

$$\begin{cases} \cos \theta_i = c'_{11} \hat{m}_{i1} + c'_{21} \hat{m}_{i2} \\ \sin \theta_i = c'_{12} \hat{m}_{i1} + c'_{22} \hat{m}_{i2} \\ X_{0i} = \frac{\sum_{j=1}^n x_{ij}}{nf} Y_{0i} \end{cases} \quad i = 1, 2, \dots, m \quad (5.35)$$

そして, (5.34) と (5.16) を利用して物体形状は求められる.

$$\begin{cases} X_j = (d'_{11} \hat{s}_{1j} + d'_{12} \hat{s}_{2j}) Y_0 \\ Y_j = (d'_{21} \hat{s}_{1j} + d'_{22} \hat{s}_{2j}) Y_0 \end{cases} \quad j = 1, 2, \dots, n \quad (5.36)$$

ここで

$$\begin{cases} d'_{11} = \frac{c'_{22}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} \\ d'_{21} = \frac{-c'_{21}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} \\ d'_{12} = \frac{-c'_{12}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} \\ d'_{22} = \frac{c'_{11}}{\begin{vmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{vmatrix}} \end{cases}$$

自由変数 Y_0 は特徴点の平均深さで, スケールファクタとして物体形状とカメラ運動の情報に残る.

5.3 カメラの移動量が多い場合の物体形状とカメラ運動の推定

5.2 では物体の大きさはカメラとの距離に比べて十分小さく, しかも画像が撮像される間におけるカメラの運動は少ない場合の物体形状とカメラ運動を推定する手法を紹介した. この節では物体の大きさは物体とカメラ間の距離に比べて十分小さく, しかしカメラの移動は大きい場合の物体形状とカメラ運動を推定する手法を紹介する.

一般的にカメラが物体との距離を一定に保ちながら移動する場合は希である. 従って, 5.2 で紹介した方法は直接に適用することができない.

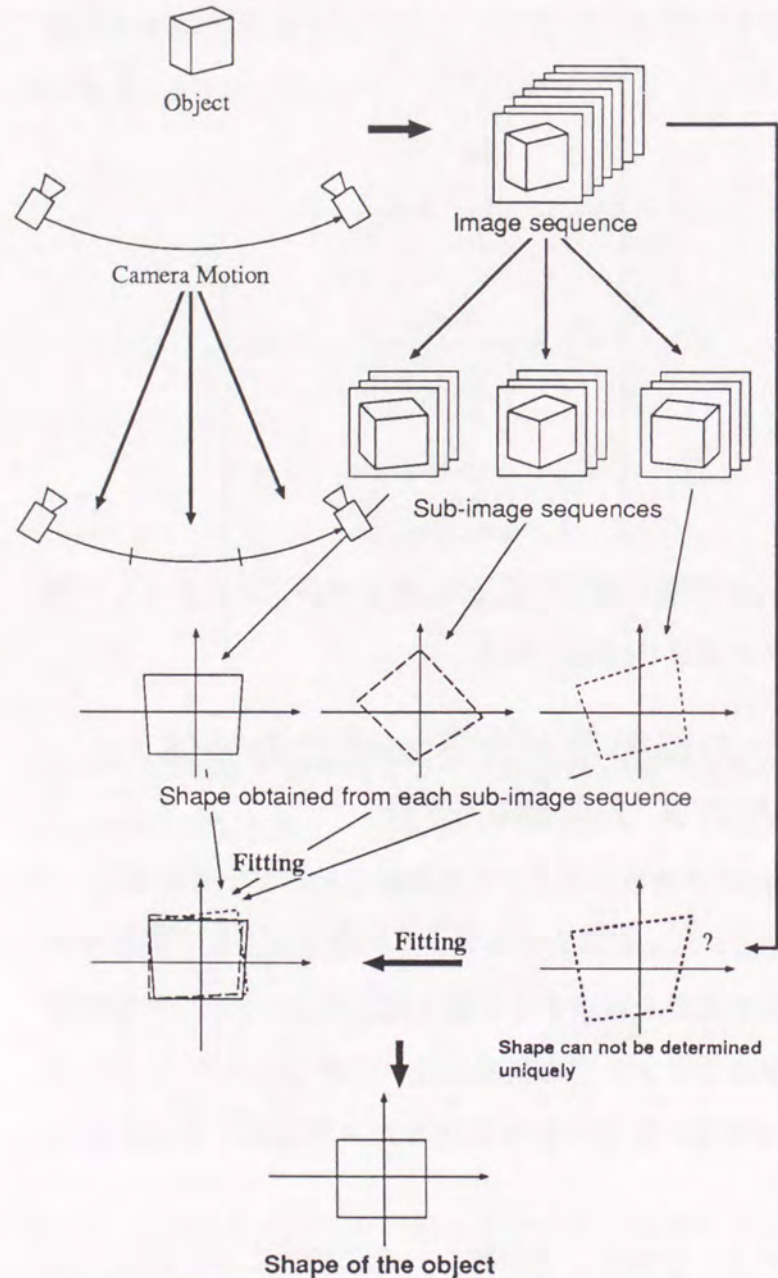


Figure 9 Acquiring the structure of scene with general camera motion.

この場合, 図9が示すように, カメラ運動の平滑性により, 連続画像を分割すれば, 各サブ画像列が撮像される間におけるカメラ運動は小さくなり, 各サブ画像列に対して5.2で紹介した方法を適用して, 物体形状とサブ画像列が撮像される間におけるカメラ運動を推定できる.

分割された連続画像は I_1, I_2, \dots, I_l とし, 各サブ画像列 I_k に5.2で紹介した方法を適用し, 求められた物体形状とカメラ運動を S_k と M_k とする.

$$S_k = \begin{pmatrix} X_{k1} & X_{k2} & \dots & X_{kn} \\ Y_{k1} & Y_{k2} & \dots & Y_{kn} \end{pmatrix} \quad (5.37)$$

$$M_k = \begin{pmatrix} \cos \theta_{k1} & \cos \theta_{k2} & \dots \\ \sin \theta_{k1} & \sin \theta_{k2} & \dots \\ X_{0k1} & X_{0k2} & \dots \end{pmatrix} \quad (5.38)$$

ここで, 我々は世界座標系の Y 軸を I_1 の最初のフレームが撮像される時のカメラの光軸と並行するように設定し, I_1 が撮像される時の物体の平均深さ Y_{01} を1とする.

S_1, S_2, \dots, S_l は同じ物体の形状情報なので, 各形状情報は同じになるはずである. 5.2で紹介したように, 画像列の最初のフレームが撮像される時のカメラの光軸の方向を物体形状を表現する座標系の Y 軸の方向としているので, S_1, S_2, \dots, S_l の参照座標系の間にある回転角が存在する. そして各サブ画像列が撮像される時の物体の平均深さ Y_{0i} は自由変数なので, S_1, S_2, \dots, S_l の座標の間にはあるスケールファクタが存在する. 従って, 次の

式が得られる.

$$\begin{cases} X_{1j} = X_{kj}kc_k - Y_{kj}ks_k & j = 1, 2, \dots, n \\ Y_{1j} = X_{kj}ks_k + Y_{kj}kc_k & k = 1, 2, \dots, l \end{cases} \quad (5.39)$$

ここで

$$\begin{cases} kc_k = \cos \alpha_k \frac{Y_{01}}{Y_{0k}} \\ ks_k = \sin \alpha_k \frac{Y_{01}}{Y_{0k}} \end{cases} \quad (5.40)$$

式の中の α_k はサブ画像列 I_k から推定されたシーンの構造とカメラ運動の参照座標系の Y 軸とサブ画像列 I_1 から推定されたシーンの構造とカメラ運動の参照座標系の Y 軸の間の角度で, Y_{01} と Y_{0k} は I_1 と I_k が撮像される時の物体の平均深さで, (X_{1j}, Y_{1j}) と (X_{kj}, Y_{kj}) は I_1 と I_k から推定された j 番目の特徴点の位置座標である.

形状情報 (S_2, S_3, \dots, S_l) を S_1 の参照座標系で表現するように, 次の変換を行なう.

$$\begin{cases} X'_{kj} = X_{kj}kc_k - Y_{kj}ks_k & j = 1, 2, \dots, n \\ Y'_{kj} = X_{kj}ks_k + Y_{kj}kc_k & k = 1, 2, \dots, l \end{cases} \quad (5.41)$$

$(X'_{1j}, Y'_{1j}), (X'_{2j}, Y'_{2j}), \dots, (X'_{kj}, Y'_{kj})$ は同じ特徴点の座標なので, その値は同じになるはずである. 故に, 次の制約条件が得られる.

$$\sum_{k=1}^l \sum_{j=1}^n ((X'_{kj} - \bar{X}'_j)^2 + (Y'_{kj} - \bar{Y}'_j)^2) \rightarrow \min \quad (5.42)$$

ここで

$$\begin{cases} \bar{X}'_j = \frac{\sum_{k=1}^l X'_{kj}}{l} \\ \bar{Y}'_j = \frac{\sum_{k=1}^l Y'_{kj}}{l} \end{cases} \quad (5.43)$$

最小自乗法を利用して(5.43)を解き, 座標変換のパラメータが求められる. これらのパラメータを利用して, (5.41)により各サブ画像列から推定された物体形状を同じ座標系に変換する.

しかし, カメラの移動が少ない場合, 連続画像から推定された物体形状の誤差は大きく信頼性は低い. 物体の形状を信頼性高く推定するために, 画像列の全体を利用する必要がある.

5.1節により, 物体の大きさはカメラとの距離に比べて十分小さい場合, 画像列を表現する行列を *singular value decomposition* のアルゴリズムを利用して分解して, (5.16)により, 次の結果が得られる.

$$\begin{cases} X_j = d_{11}\hat{s}_{1j} + d_{12}\hat{s}_{2j} \\ Y_j = d_{21}\hat{s}_{1j} + d_{22}\hat{s}_{2j} \end{cases} \quad j = 1, 2, \dots, n$$

ここで, (X_j, Y_j) は物体の特徴点の位置で, $d_{11}, d_{12}, d_{21}, d_{22}$ は自由変数である. $\hat{s}_{1j}, \hat{s}_{2j}$ の値は画像列から求められる.

物体の形状情報を得るために, 変数 $d_{11}, d_{12}, d_{21}, d_{22}$ を決定しなければならない. (X_j, Y_j) は (X_{kj}, Y_{kj}) $k = 1, 2, \dots, l$ と同じように j 番目の物体の特徴点の位置座標なので, (X_j, Y_j) は (X_{kj}, Y_{kj}) $k = 1, 2, \dots, l$ と同じ値をとるはずである. 故に次の式

が得られる.

$$\begin{cases} \sum_{j=1}^n \sum_{k=1}^l (d_{11}s'_{1j} + d_{12}s'_{2j} - X'_{kj})^2 \rightarrow \min \\ \sum_{j=1}^n \sum_{k=1}^l (d_{21}s'_{1j} + d_{22}s'_{2j} - Y'_{kj})^2 \rightarrow \min \end{cases} \quad (5.44)$$

最小自乗法を利用して d_{11} , d_{12} , d_{21} , d_{22} の値を決定して, (5.16) を利用して (X_j, Y_j) の値を決定する. d_{11} , d_{12} , d_{21} , d_{22} 及び (5.17) の方程式を用いて c_{11} , c_{12} , c_{21} , c_{22} を計算して, (5.15) と (5.8) からカメラ運動を次のように求める.

$$\begin{cases} Y_{0i} = \frac{1}{\sqrt{(c_{11}\hat{m}_{i1} + c_{21}\hat{m}_{i2})^2 + (c_{12}\hat{m}_{i1} + c_{22}\hat{m}_{i2})^2}} \\ \cos \theta_i = (c_{11}\hat{m}_{i1} + c_{21}\hat{m}_{i2})Y_{0i} \\ \sin \theta_i = (c_{12}\hat{m}_{i1} + c_{22}\hat{m}_{i2})Y_{0i} \\ X_{0i} = \frac{\sum_{j=1}^n x_{ij}}{nf} Y_{0i} \end{cases} \quad (5.45)$$

6

深さ範囲の広いシーンの構造の獲得

シーンの深さ範囲が広いとき、シーンからカメラへの投影は平行射影で近似できなくなり、前章で紹介した方法は直接に適用できない。もし、シーンは何らかの方法で深さがほぼ一定のグループに分割できれば、各グループのシーンの深さ範囲は狭くなり、各グループのシーンに対してカメラへの投影は平行射影で近似できて、前章で紹介した方法を用いてシーンの構造を推定することができる。これらのシーンの構造の情報を融合して、シーン全体の構造が復元できる。

深さ範囲が広い一般的なシーンに対して、前章で紹介した方法を直接に適用すると、算定したシーンの構造には歪みが含まれるようになる。しかし、特徴点の大まかな配置を知ることができる。図10が示すように、この情報を利用してシーンを分割し、画像列もシーンの分割に応じていくつかのサブ画像列に分割する。各グループのシーンの構造とカメラ運動をサブ画像列から推定する。

全てのサブ画像列は同じカメラで撮像したものであるため、サブ画像列から推定されたカメラ運動は同じになるはずである。この条件を用いて、各グループのシーンの構造の間の相対位置を算定し、全てのグループのシーンを一つに融合して、シーン全体の構造を復元する。

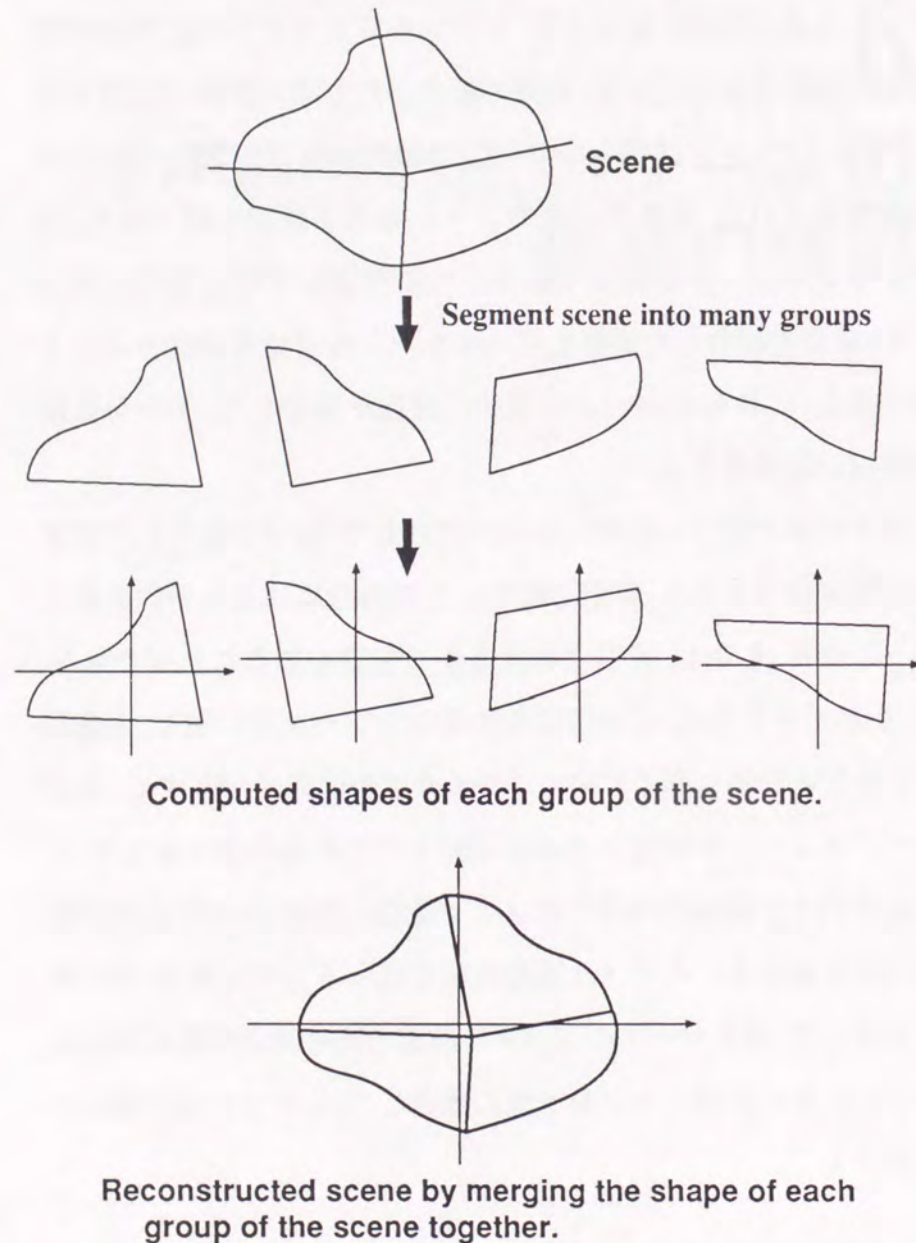


Figure 10 Acquiring the structure of a general scene.

6.1 シーンの分割

第5章で紹介した方法を用いて、深さ範囲が広い一般的なシーンの構造を推定するために、シーンを大きさがカメラとの距離に比べて十分小さいグループに分割することは必要である。しかし、この分割にはシーンの構造の情報、即ちシーンの中の特徴点の配置の情報が必要である。

深さ範囲が狭いと仮定して、前章で紹介した方法を用いてシーンの構造の情報が獲得できるが、その情報には歪みが含まれる。しかし、特徴点の大まかな配置を知ることができるので、この情報を利用してシーンが分割できる。シーンの深さ範囲を狭くすることは分割の目的なので、シーンの構造情報の歪みによる分割した後の各グループのシーンの大きさには少々ばらつきがあっても支障はない。

シーンの分割に対応して、一つのサブ画像列には一つのグループのシーンの特徴点だけを含むように画像列を分割する。そして、前章で紹介した方法を各サブ画像列に適用して、各グループのシーンの構造とカメラ運動を推定する。

6.2 分割されたシーンの融合による環境構造の復元

シーンの深さ範囲を無視して、第5章で紹介した手法を連続画像に適用し、シーンの構造を求め、6.1節で紹介した方法を用いてシーンを分割し、それと同時に一つのサブ画像列には一つ

のグループのシーンの特徴点だけを含むように画像列を分割する。分割された画像列は I_1, I_2, \dots, I_l とし、各サブ画像列に対して、第5章で紹介した方法を適用して、各グループのシーンの構造 S_1, S_2, \dots, S_l とカメラ運動 M_1, M_2, \dots, M_l を推定する。ここで

$$S_k = \begin{pmatrix} X_{k1} & X_{k2} & \dots \\ Y_{k1} & Y_{k2} & \dots \end{pmatrix}$$

$$M_k = \begin{pmatrix} X_{0k1} & X_{0k2} & \dots & X_{0km} \\ Y_{0k1} & Y_{0k2} & \dots & Y_{0km} \\ \theta_{k1} & \theta_{k2} & \dots & \theta_{km} \end{pmatrix}$$

第5章で紹介した方法はシーンの構造とカメラ運動を推定するとき、世界座標系の原点はシーンの重心に選定され、画像列の最初フレームが撮像されるときシーンの平均深さは座標系の長さの単位としているので、各サブ画像列から推定されたシーンの構造とカメラ運動は参照する世界座標系の中に座標原点の位置の差とスケールの差が存在する。そして各座標系の座標軸の方向にもある回転角が存在する。

我々はまず各グループのシーンの構造の情報とカメラ運動の参照座標系の軸の方向を統一させる。 (M_1, S_1) の参照座標系を統一の基準とする。 (M_k, S_k) の参照座標系と (M_1, S_1) の参照座標系との角度を $d\theta_k$ とし、 I_k から求めたカメラの回転運動のパラメータは統一の参照座標系で次のように表現できる。

$$\theta'_{ki} = \theta_{ki} + d\theta_k \quad (6.1)$$

式の中の θ_{ki} は I_k から求められた i 番目フレームの画像が撮像さ

れるときのカメラの回転角度で、 θ'_{ki} は変換を行った後のカメラの回転角度である。

各サブ画像列は同じカメラから撮像されたものなので、サブ画像列から得られたカメラ運動は同じになるはずである。従って、全ての k に対して θ'_{ki} は同じ値になるはずなので、次の制約条件が得られる。

$$\sum_{k=0}^l \left(\sum_{i=0}^m (\theta_{ki} - \bar{\theta}_i)^2 \right) \rightarrow \min \quad (6.2)$$

ここで

$$\bar{\theta}_i = \frac{\sum_{k=0}^l \theta_{ki}}{l} \quad (6.3)$$

最小自乗法を利用して (6.1) と (6.2) から $d\theta_1, d\theta_2, \dots, d\theta_l$ を算定して、各サブ画像列 I_k から推定されたカメラの回転運動は (6.1) を用いて統一参照座標系への変換を行う。

3章で説明したように、カメラの並進運動 (X_{0i}, Y_{0i}) はカメラ座標系での世界座標系の原点の座標で表現される。後の処理の便利のために、カメラ運動を世界座標系での座標で表現するように変換する。

$$\begin{cases} X_{cki} = -\cos \theta'_{ki} X_{0ki} + \sin \theta'_{ki} Y_{0ki} \\ Y_{cki} = -\sin \theta'_{ki} X_{0ki} - \cos \theta'_{ki} Y_{0ki} \end{cases} \quad (6.4)$$

(M_1, S_1) の参照座標系と (S_k, S_k) の参照座標系の原点座標の位置の差とスケールファクタの差を dX_k, dY_k, k_k とし、 $(M_2, S_2), (M_3, S_3), \dots, (M_l, S_l)$ の参照座標系を (M_1, S_1) の参照座標系に

統一させ、次の式を用いて各サブ画像列から求められたカメラの並進移動を (M_1, S_1) の座標系への変換を行う。

$$\begin{cases} X'_{cki} = X_{cki}k_k + dX_k \\ Y'_{cki} = Y_{cki}k_k + dY_k \end{cases} \quad (6.5)$$

式の中の X_{cki}, Y_{cki} は I_k から求められた i 番目のフレームの画像が撮像されるときのカメラの位置の座標で、 (X'_{cki}, Y'_{cki}) は座標変換を行った後のカメラの位置の座標である。

全てのサブ画像列は同じカメラから撮像されたものなので、全ての k に対して (X'_{cki}, Y'_{cki}) は同じのはずなので、次の式が得られる。

$$\sum_{k=1}^l \sum_{i=1}^m ((X'_{cki} - \overline{X'_{ci}})^2 + (Y'_{cki} - \overline{Y'_{ci}})^2) \rightarrow \min \quad (6.6)$$

ここで

$$\begin{cases} \overline{X'_{ci}} = \frac{\sum_{k=1}^l X'_{cki}}{l} \\ \overline{Y'_{ci}} = \frac{\sum_{k=1}^l Y'_{cki}}{l} \end{cases} \quad (6.7)$$

最小自乗法を利用して、(6.5) と (6.6) から k_k, dX_k, dY_k ($k = 1, 2, \dots, l$) が求められ、(6.5) を利用してカメラの並進移動を (M_1, S_1) の参照座標系に変換することができる。

求められた各参照座標系間の回転角度 $d\theta_1, d\theta_2, \dots, d\theta_l$ とスケールファクタ k_1, k_2, \dots, k_l 及び座標原点の位置の差 $(dX_1, dY_1), (dX_2, dY_2), \dots, (dX_l, dY_l)$ のを利用して、各グループのシーンの

構造情報 S_1, S_2, \dots, S_l に対して (M_1, S_1) の参照座標系への変換を行なう。

$$\begin{cases} X'_{kj} = k_k(X_{kj} \cos d\theta_k - Y_{kj} \sin d\theta_k) + dX_k \\ Y'_{kj} = k_k(X_{kj} \sin d\theta_k + Y_{kj} \cos d\theta_k) + dY_k \end{cases} \quad (6.8)$$

式の中の (X_{kj}, Y_{kj}) は k 番目のグループのシーンの j 番目の特徴点の位置座標で、 (X'_{kj}, Y'_{kj}) は座標変換を行った後のものである。

こうすると、全ての特徴点の座標は同じ座標系を参照することになり、これらの特徴点の位置の情報はシーンの構造を示すのである。

7

ニュートン法による シーンの構造と カメラ運動の推定

静止環境においてカメラを移動させ、撮像された連続画像から特徴点を抽出し、フレーム間の対応づけを行なった後、6章で紹介した方法を用いてシーンの構造とカメラ運動を推定することができる。しかし、6章の方法を用いてシーンの構造とカメラ運動を推定する時、多種の近似処理を行うので、推定されたシーンの構造とカメラ運動には歪みや誤差が含まれると考えられる。この章では、6章で紹介した方法で推定されたシーンの構造とカメラ運動の情報をシーンからカメラへの投影の方程式の解の初期値とし、ニュートン法を利用して、カメラの投影の非線形方程式を解くことにより、シーンの構造とカメラ運動を推定する方法を紹介する。

3章で説明したように、シーンからカメラへの投影は次の式で表現することができる。

$$x = f \frac{X \cos \theta + Y \sin \theta + X_0}{-X \sin \theta + Y \cos \theta + Y_0} \quad (7.1)$$

シーンの特徴点の数は n とし、連続画像に m フレームがあるとする。この場合、次のような、 $m \times n$ 個の方程式が得られる。

$$x_{ij} = f \frac{X_j \cos \theta_i + Y_j \sin \theta_i + X_{0i}}{-X_j \sin \theta_i + Y_j \cos \theta_i + Y_{0i}} \quad \begin{cases} i = 1, 2, \dots, m \\ j = 1, 2, \dots, n \end{cases} \quad (7.2)$$

シーンの構造パラメータ $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ とカメラの回転運動パラメータ $\theta_1, \theta_2, \dots, \theta_m$ 及びカメラの並進運動パラメータ $(X_{01}, Y_{01}), (X_{02}, Y_{02}), \dots, (X_{0m}, Y_{0m})$ はこれらの方程式を解くことによって求められる。6章で紹介した手法を用いてこれらの方程式の近似解 $(X'_1, Y'_1), (X'_2, Y'_2), \dots, (X'_n, Y'_n), \theta'_1,$

$\theta'_2, \dots, \theta'_m, (X'_{01}, Y'_{01}), (X'_{02}, Y'_{02}), \dots, (X'_{0m}, Y'_{0m})$ を求めることができる。

(7.2) を微分して、2 次以上の項を省略して次の近似式が得られる。

$$\frac{\partial x_{ij}}{\partial \theta_i} \Delta \theta_i + \frac{\partial x_{ij}}{\partial X_j} \Delta X_j + \frac{\partial x_{ij}}{\partial Y_j} \Delta Y_j + \frac{\partial x_{ij}}{\partial X_{0i}} \Delta X_{0i} + \frac{\partial x_{ij}}{\partial Y_{0i}} \Delta Y_{0i} \approx \Delta x_{ij} \quad (7.3)$$

ここで

$$\left\{ \begin{array}{l} \frac{\partial x_{ij}}{\partial \theta_i} = f \frac{V}{(-\sin \theta'_i X'_j + \cos \theta'_i Y'_j + Y'_{0i})^2} \\ \frac{\partial x_{ij}}{\partial X_{0i}} = f \frac{1}{-\sin \theta'_i X'_j + \cos \theta'_i Y'_j + Y'_{0i}} \\ \frac{\partial x_{ij}}{\partial Y_{0i}} = f \frac{-(\cos \theta'_i X'_j + \sin \theta'_i Y'_j + X'_{0i})}{(-\sin \theta'_i X'_j + \cos \theta'_i Y'_j + Y'_{0i})^2} \\ \frac{\partial x_{ij}}{\partial X_j} = f \frac{Y'_j + \sin \theta'_i X'_{0i} + \cos \theta'_i Y'_{0i}}{(-\sin \theta'_i X'_j + \cos \theta'_i Y'_j + Y'_{0i})^2} \\ \frac{\partial x_{ij}}{\partial Y_j} = f \frac{-X'_j - \cos \theta'_i X'_{0i} + \sin \theta'_i Y'_{0i}}{(-\sin \theta'_i X'_j + \cos \theta'_i Y'_j + Y'_{0i})^2} \\ \Delta \theta_i = \theta_i - \theta'_i \\ \Delta X_{0i} = X_{0i} - X'_{0i} \\ \Delta Y_{0i} = Y_{0i} - Y'_{0i} \\ \Delta X_j = X_j - X'_j \\ \Delta Y_j = Y_j - Y'_j \\ \Delta x_{ij} = x_{ij} - f \frac{X'_j \cos \theta'_i + Y'_j \sin \theta'_i + X'_{0i}}{-X'_j \sin \theta'_i + Y'_j \cos \theta'_i + Y'_{0i}} \end{array} \right. \quad (7.4)$$

ただし、

$$V = X_j'^2 + Y_j'^2 + (\cos \theta'_i X'_j + \sin \theta'_i Y'_j) X'_{0i} - (\sin \theta'_i X'_j - \cos \theta'_i Y'_j) Y'_{0i}$$

(7.3) が示めす方程式の数は $m \times n$ で、未知数は $\Delta \theta_i, \Delta X_{0i}, \Delta Y_{0i}, \Delta X_j, \Delta Y_j$ で、その数は $3 \times m + 2 \times n$ である。物体の特徴点及び画像列のフレームは多数ある場合、 $m \times n$ は $3 \times m + 2 \times n$ より大きくなり、方程式の数は未知数の数より多くなる。最小自乗法を利用して (7.3) を解き、(7.4) から得られた次の式を用いて特徴点の位置とカメラの回転、並進運動のパラメータを更新する。

$$\left\{ \begin{array}{l} \theta_i = \theta'_i + \Delta \theta_i \\ X_{0i} = X'_{0i} + \Delta X_{0i} \\ Y_{0i} = Y'_{0i} + \Delta Y_{0i} \\ X_j = X'_j + \Delta X_j \\ Y_j = Y'_j + \Delta Y_j \end{array} \right. \quad (7.5)$$

$\Delta \theta_i, \Delta X_{0i}, \Delta Y_{0i}, \Delta X_j, \Delta Y_j$ の絶対値が十分小さくなるまで、(7.3) 及び (7.5) を利用して $\theta_i, X_{0i}, Y_{0i}, X_j, Y_j$ を更新し続けることによって、シーンの構造とカメラ運動情報を正確に推定することができる。

8

実験と討論

本論文では連続画像から動物体の抽出,そして環境構造とカメラ運動の推定の新しい手法を紹介した.この章では,これらの手法の有効性を検証するための実験とその結果を紹介する.

まず第4章で紹介した動物体を抽出する手法を検証するために,次の実験を行った.実験環境は図11のように設定する.

この環境は2次元の水平面で,環境には $\{-1.0 \leq x \leq 0.75, 4.0 \leq y \leq 7.75\}$ の範囲内に分布している16個の静止の正方形があり,物体の特徴点は各正方形の頂点である.そして環境の中に1つの 0.4×0.2 のサイズの長方形の動物体があり,この移動物体の特徴点は長方形の各頂点である.移動物体は

$$x^2 + \frac{(y-3)^2}{0.5^2} = 1$$

の楕円軌道に沿って移動する.

カメラは水平面の中で,

$$x^2 + (y-5)^2 = 10^2$$

の円軌道に沿って $(0,0)$ 点を注視しながら移動する.カメラで撮像される画像は1次元のもので,画像の解像度は512画素である.そしてカメラの焦点距離は1024画素に相当する.カメラは45度を移動しながら200枚の画像を撮像する.撮像するとき,物体の特徴点はいつも見えるとし,Occlusionは存在しないと仮定する.実験にサンプリングノイズを含んだ合成画像を利用した.

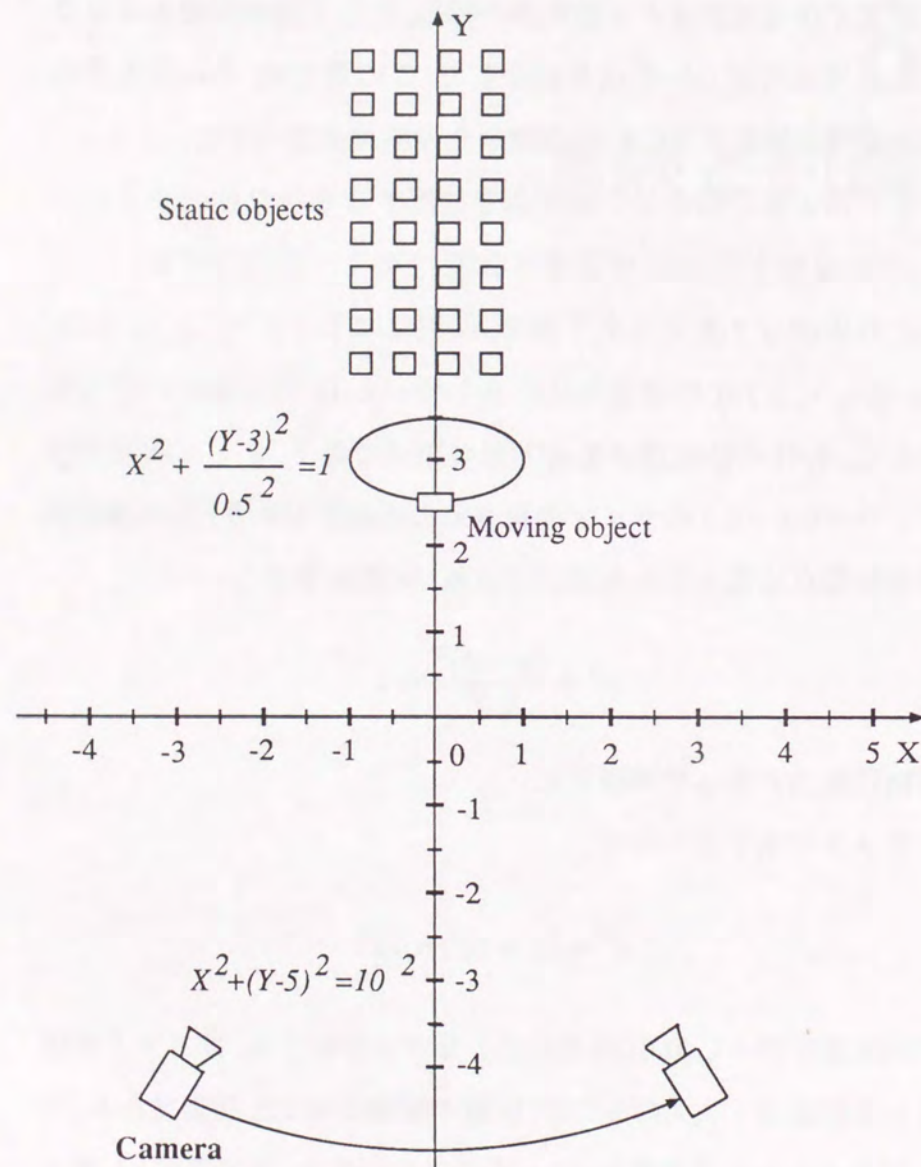


Figure 11 The setup of the environment of the experiment 1.

第4章で紹介した方法を用いて、動物体の運動が生じる画面での動きをノイズと見なし、SVDに基づくノイズを除去する方法を

用いて画像ノイズを除去する。ノイズが除去された画像と原画像の物体の特徴点の画面座標の差を計算し、各物体の特徴点に対して、全てフレームの画像の上の誤差の2乗の和を計算して、ヒストグラムを構築する。図12はその結果を示す。横軸は特徴点の番号で、縦軸は各特徴点の誤差の2乗の和であり、その中の最大の値に対するパーセンテージで表している。図12には最後の4個の点は移動物体の特徴点に対応するもので、それ以外は静止物体に対応するものである。この結果から分かるように、動物体に対応する誤差の値は静止物体に対応するものに比べてはるかに大きいので、このヒストグラムを用いて動物体と静止物体を分離することができる。

Figure 12. The sum of the square of the error of each feature point.



環境構造とカメラ運動を推定する方法を検証するために次の実験を行った. 実験環境は図13のように設定する.

この環境は2次元の水平面で, 環境には9個の静止の正方形があり, 物体の特徴点は各正方形の辺に分布している.

カメラは水平面の中で,

$$\frac{x^2}{7.5^2} + \frac{y^2}{5.0^2} = 1$$

の楕円軌道に沿って(0,0)点を注視しながら移動する. カメラで撮像される画像は1次元のもので, 画像の解像度は512画素である. そしてカメラの焦点距離は1024画素に相当する. カメラは148度を移動しながら201枚の画像を撮像する. 撮像するとき, 物体の特徴点はいつも見えるとし, Occlusionは存在しないと仮定する. 実験にサンプリングノイズを含んだ合成画像を利用した.

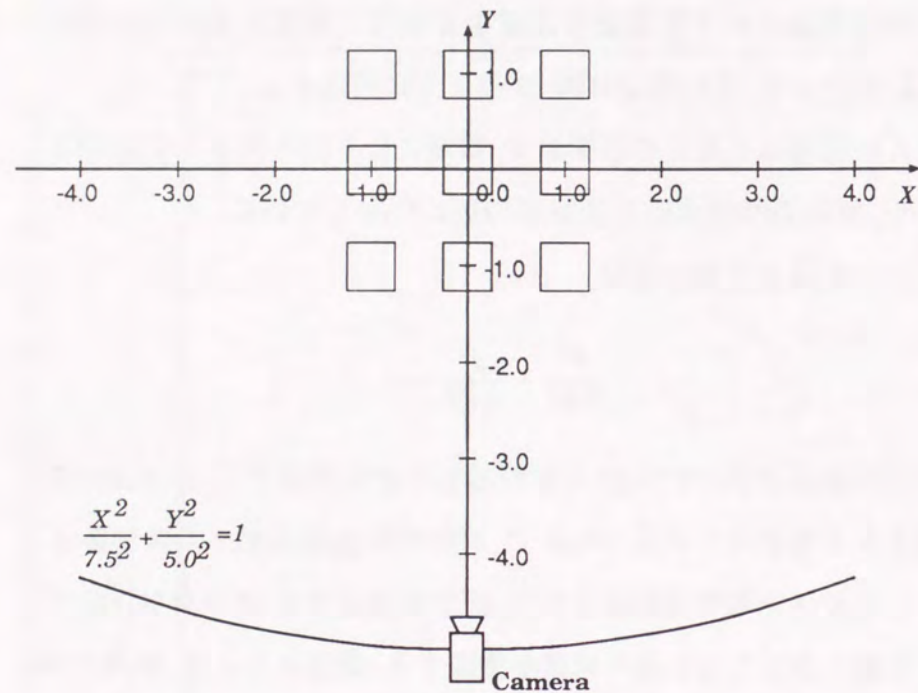


Figure 13 The setup of the environment of the experiment 2.

まず物体の特徴点の深さ範囲が狭いと仮定して、第5章で紹介した深さ範囲が狭い場合の物体構造とカメラ運動を推定する方法を用いてシーンの構造を推定する。図14は推定されたシーンの構造の結果を示す。

図14から分かるように、この結果には強い歪みが含まれ、物体の形状が歪んでいる。しかし、物体の大体の形状及び物体間の位置関係はこの結果から知ることができる。

第6章で紹介したように、図14で示したシーンの構造の情報を利用してシーンを分割し、画像列を分割されたシーンのグループ

の特徴点だけを含むサブ画像列に分割する。各グループのシーンの構造とカメラ運動をサブ画像列から推定し、そして各サブ画像列から推定されたカメラ運動は不変の条件を用いて、各グループのシーンの構造の情報を融合してシーン全体の構造を復元する。図15は推定されたシーンの構造とカメラ運動の結果を示す。

図15から分かるように、この結果にはやはり歪みが含まれている。しかし、図14で示した結果に比べて、歪みは少なく、物体の形状の変形は少ない。図15が示した物体の形状とカメラ運動のスケールは統一されていない。

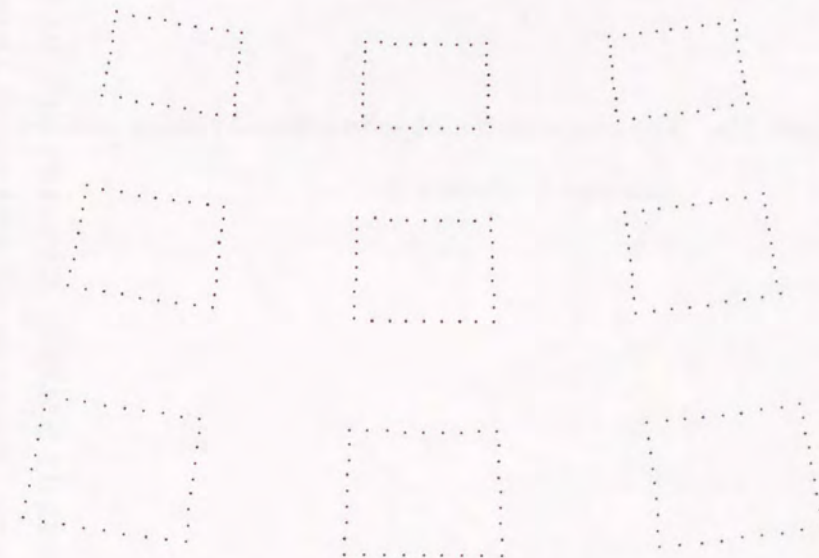


Figure 14. The shape of the objects obtained using the method described in chapter 5.

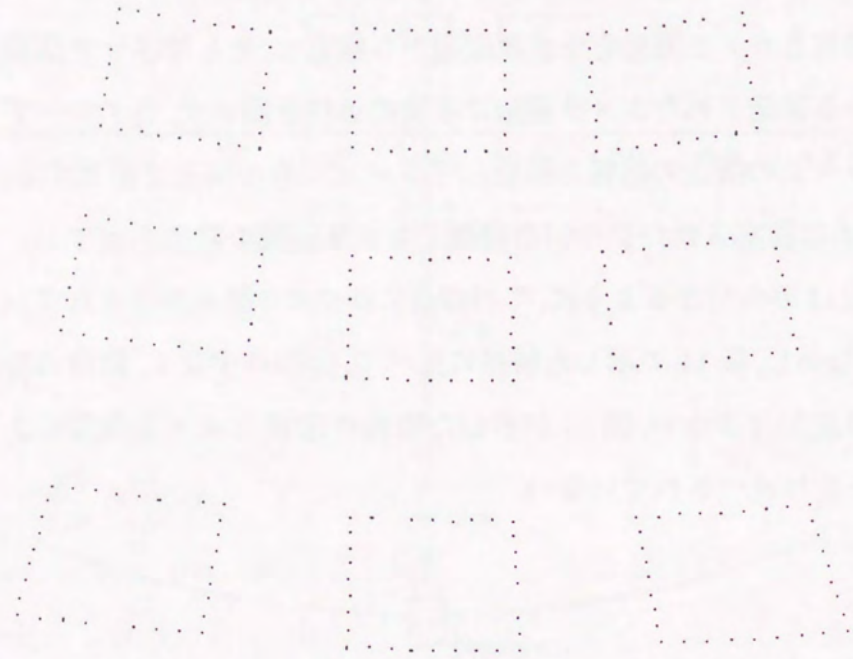


Figure 15a. The shape of the objects obtained using method describe in chapter 6.

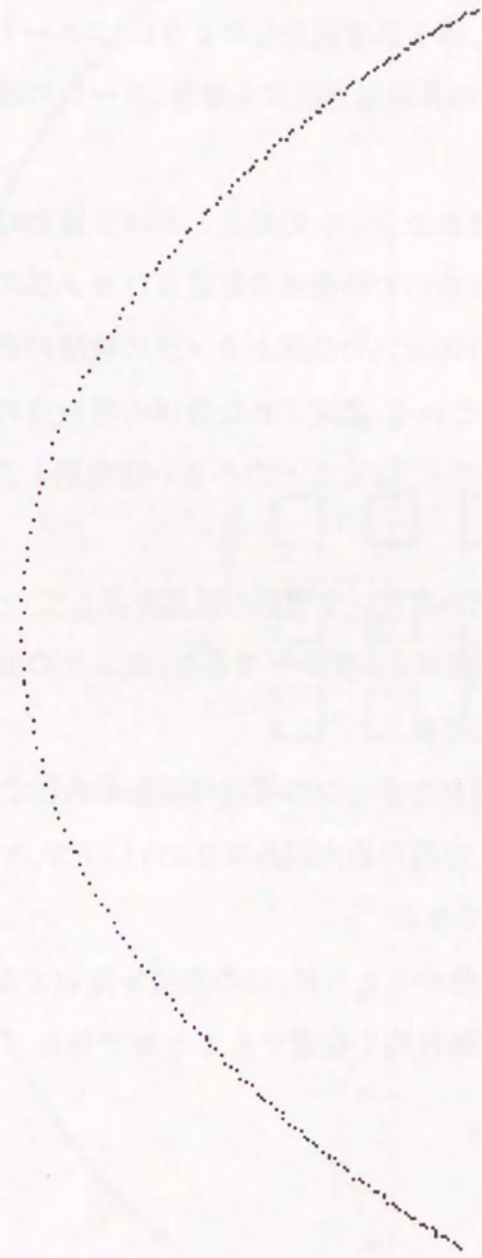


Figure 15b The trajectory of the camera. Notice that it is obviously different from the theoretical one.

図15で示した結果をシーンからカメラへの投影の方程式の解の初期値として, 第7章で紹介したように, ニュートン法を利用してカメラの投影の非線形方程式を解き, シーンの構造とカメラ運動を推定する.

図16は推定されたシーンの構造とカメラ運動の結果を示している. 推定された物体の特徴点の位置及びカメラの軌跡は黒い点で表示し, 物体の特徴点の位置とカメラの軌跡の理論値は小さい四角形で表示している. 推定された物体の特徴点の位置とカメラ運動の誤差が小さく, ほとんどの小さい四角形と黒い点は重なっている.

図17は推定されたカメラ運動の誤差を示している. カメラの回転運動の最大誤差は0.6度以下であり, カメラの並進運動の最大誤差は0.06以下である.

図18は推定されたシーンの構造の誤差を示している. 物体の特徴点の位置のX方向の最大誤差は0.005以下で, Y方向の最大誤差は0.006以下である.

この結果から分かるように, この方法を用いてシーンの構造とカメラ運動を信頼性高く推定することができる.

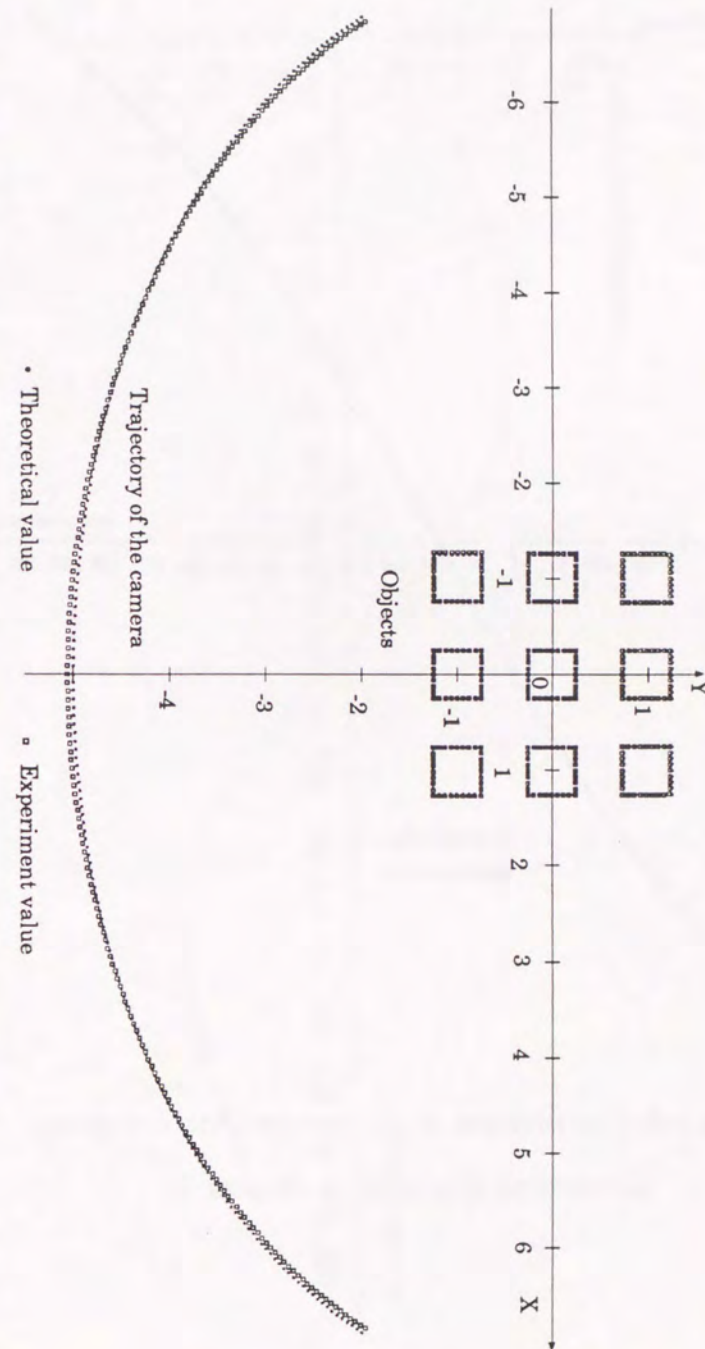


Figure 16a The translation for the camera and the shape of the objects obtained using the method described in chapter 7.

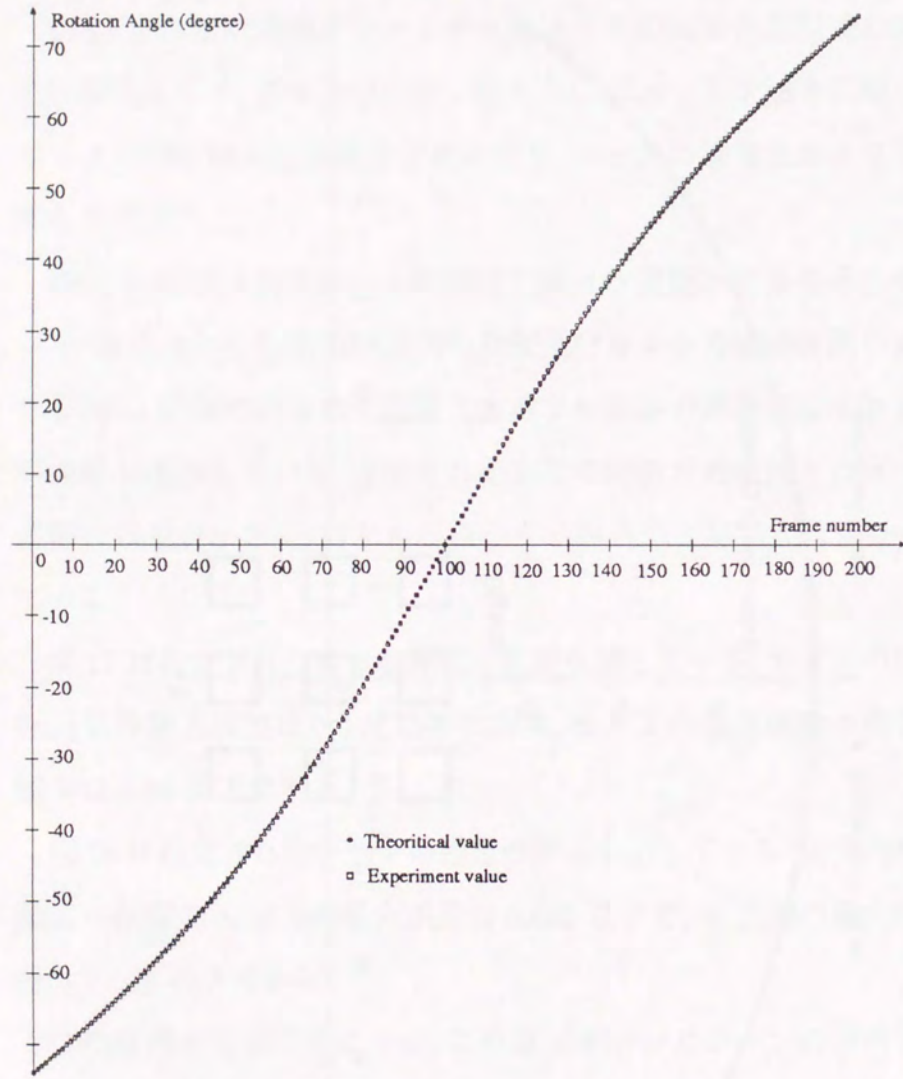


Figure 16b The rotation of the camera obtained using the method described in chapter 7.

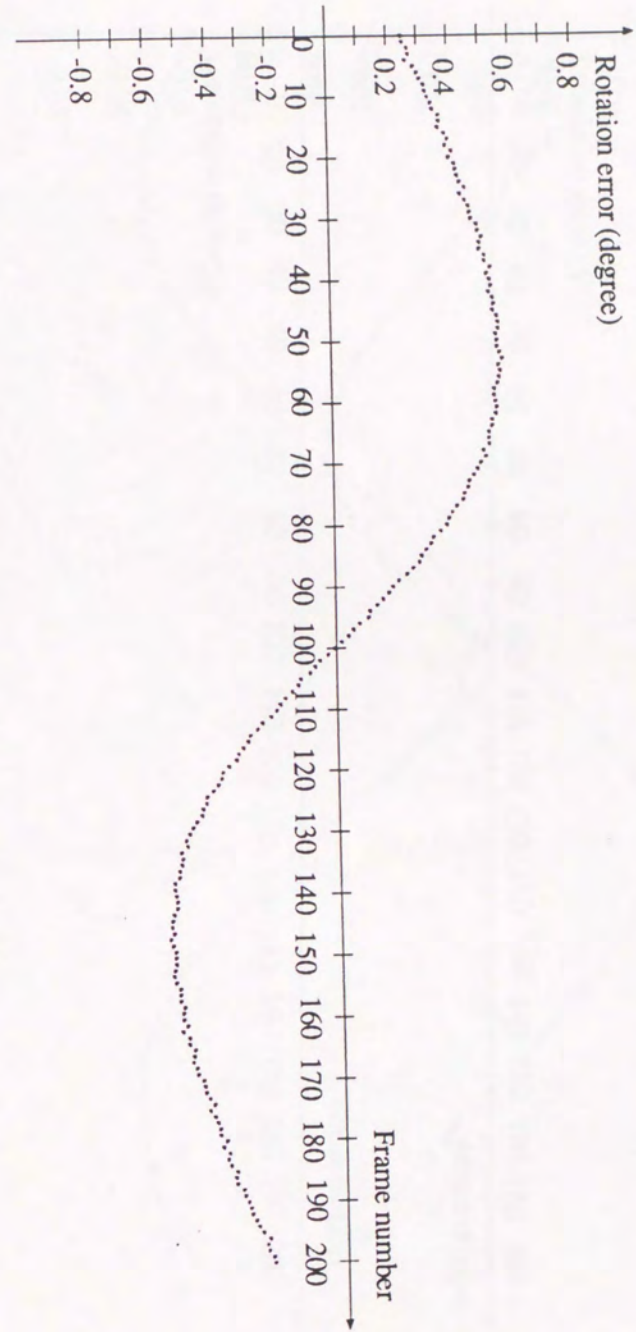


Figure 17a. The rotation error of the camera motion.

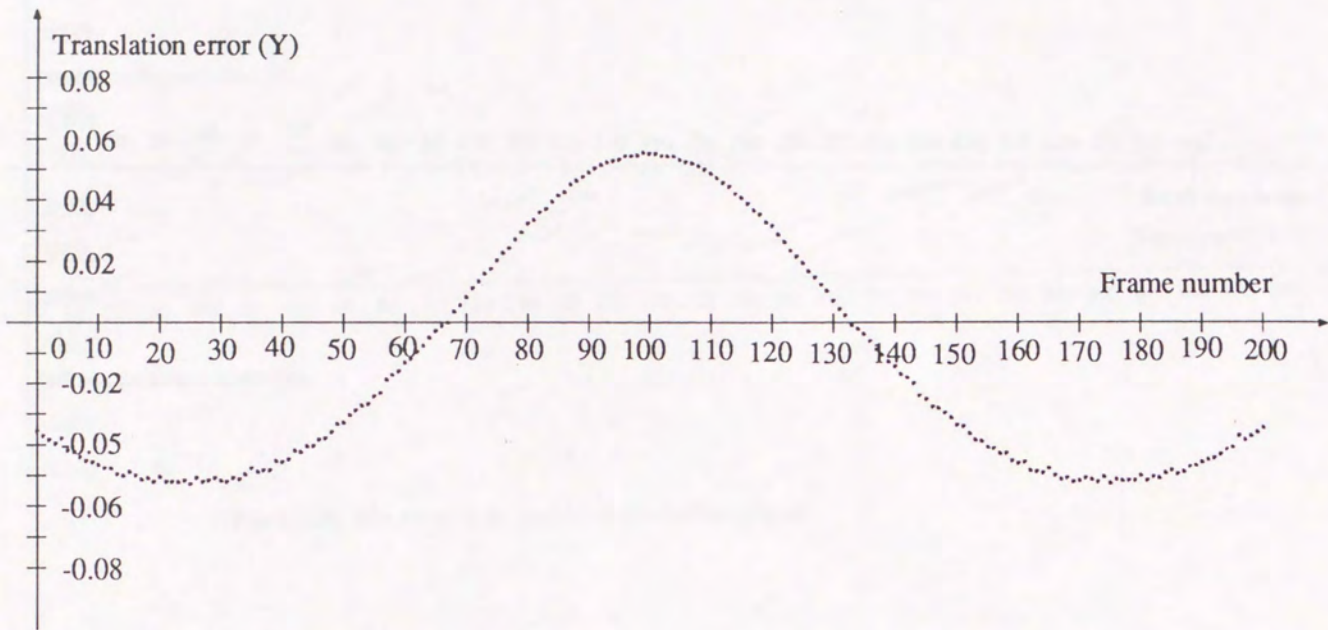


Figure 17c. The translation error in Y direction of the camera motion.

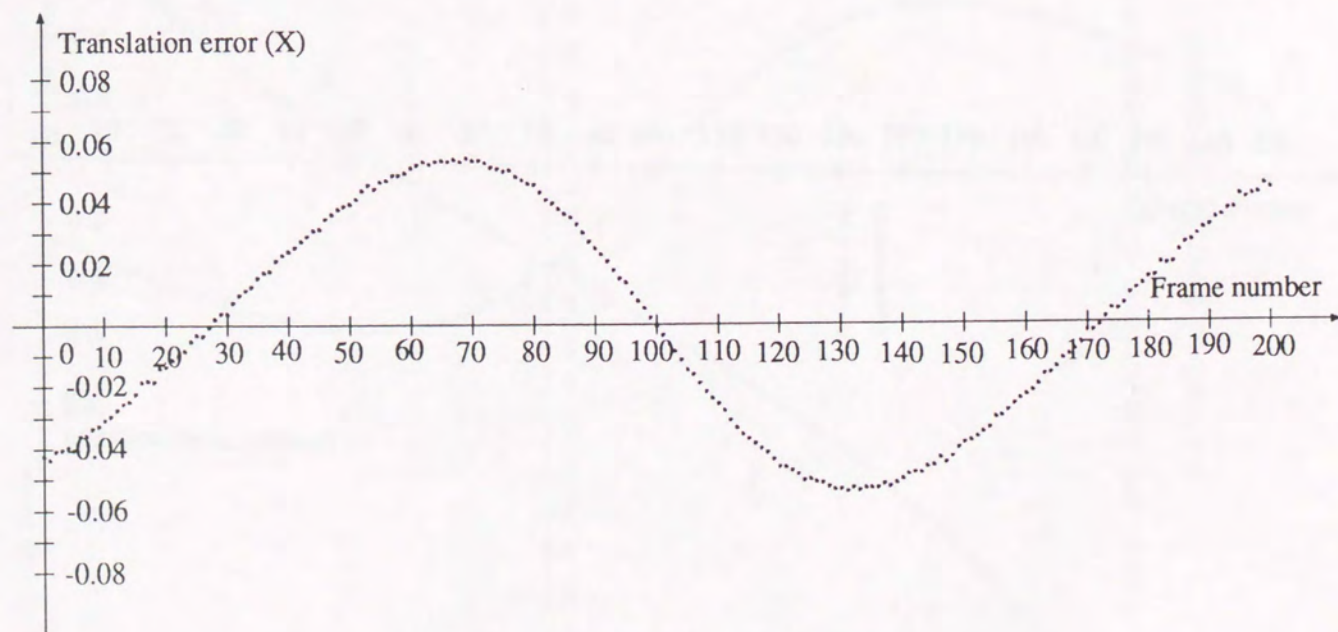


Figure 17b. The translation error in X direction of the camera motion.

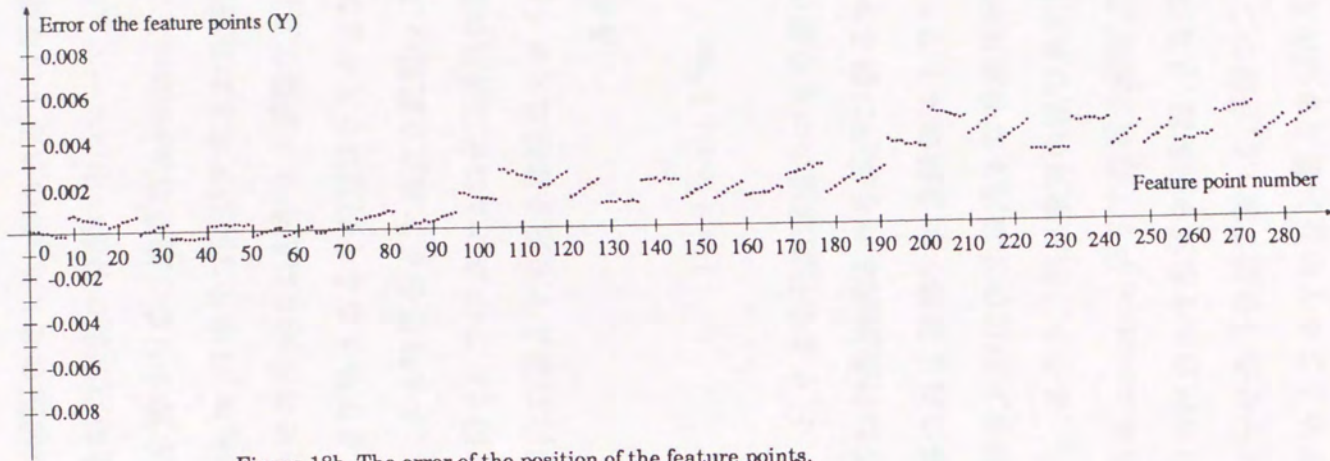


Figure 18b. The error of the position of the feature points.

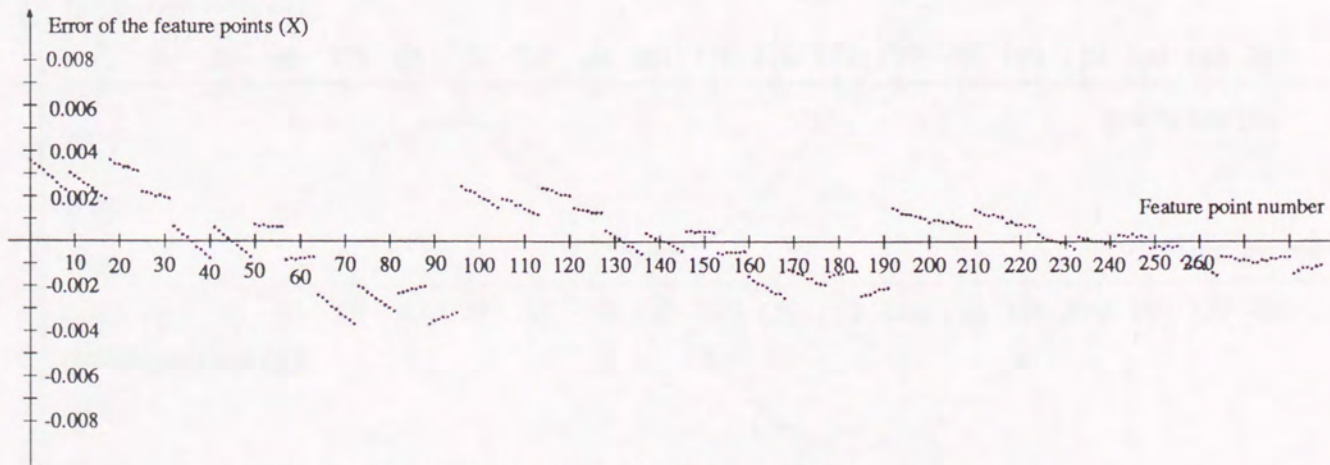


Figure 18a. The error of the position of the feature points (X).

本研究では議論を簡単にするために、カメラの運動は水平面に限定していると仮定し、シーンの中の物体の特徴点は全てある水平面上にあり、そしてocclusionは存在しないと仮定している。

現実では、物体の特徴点は3次元空間に分布し、カメラ運動も3次元空間に行く。本研究で提案した手法は3次元のシーンとカメラ運動の場合で適用できるように拡張することができる。しかし、拡張する時に次の新しい課題を克服する必要がある。

2次元のカメラ運動の回転成分は1つなので、5.2節で説明したように、シーンの構造とカメラ運動を決定する時に利用される拘束条件の式は単純である。

$$\cos^2 \theta + \sin^2 \theta = 1$$

3次元のカメラ運動の場合、3つの回転成分があり、シーンの構造とカメラ運動を決定する時に利用される拘束条件はこの3つの回転角度のサインとコサインの組合せで構成された複雑な式になる。結果的に、この拘束条件は4次以上の方程式を解く問題になり、この方程式の解法及び解の判定は難しくなる。

実際の環境において、画像を撮像する時occlusionの現象が起こる。Occlusionの現象により、画像列を記述する行列には未知の値を持つ要素が現れる。この場合、5.1節で紹介したシーンの構造とカメラ運動を分離する方法を直接に適用することができなくなる。従って、5.1節の手法に代る方法を見つける必要がある。

9 結論

本論文は structure from motion についての研究を述べたものである。本論文では、まず従来の structure from motion の分野においてほとんど研究されなかった移動物体が存在する場合の環境構造の推定の問題について研究し、*singular value decomposition* という数学の道具を利用して、静止物体と移動物体を分離する手法を提案し、その原理について述べた。

そして、従来の研究において、連続画像から深さ範囲の広いシーンの構造とカメラ運動の推定という未解決の問題に対して、まず深さ範囲が狭い場合に適用できるシーンの構造とカメラ運動を推定する方法を考案した。そして、この方法を用いて深さ範囲の広い一般的なシーンに適用して、環境の大まかな構造を獲得し、この情報を利用してシーンを同程度の深さ範囲を持つグループに分割し、画像列を各グループのシーンの特徴点だけを含むサブ画像列に分割する。各サブ画像列に対して、上記の手法を利用して各グループの形状とカメラ運動を推定する。そして、算定されたカメラ運動は不変という条件を用いて各グループの形状情報を融合してシーン全体の構造を復元する。推定されたシーンの構造とカメラ運動をカメラへの投影の方程式の解の初期値として、ニュートン法を利用してシーンの構造とカメラ運動を正確に推定する。

実験結果が示したように、本論文で提案した方法を用いて静止物体と移動物体を分離することができ、そして深さ範囲が広い一般的なシーンにおいて連続画像から環境構造とカメラ運動は信

頼性高く推定できる. 本研究は structure from motion の研究の進展を与えたものと確信する.

謝辞

この博士論文は多くの人々の助言や協力を得て執筆することができたものである。末筆ながら、ここに心から感謝の意を表します。

まず最初に、本論文選考の主査であり私の指導教官である辻三郎教授に深く感謝いたします。修士課程に入学して現在に至るまで、辻教授は常に貴重な御指導と励ましの言葉を頂き、本論文の遂行に当たっては懇切な御討論、創造溢れる御指導、御鞭撻を賜り、心から深く感謝致します。

本論文に関して適切な御指摘、御指導を頂いた井口征士教授と谷内田正彦教授に感謝します。

今井正和講師、徐剛助手、早瀬敏夫技官には研究室での活動において常に励まし、助けて頂いたことを感謝します。

笠井健教授にはいろいろの方面において助けて頂いたことを感謝します。鄭絳宇氏、石黒浩氏、李仕剛氏らには、共通の分野で多くの討論を持てたことを感謝します。高畑実君、加藤浩仁君には計算機システムを中心に多くの助けを頂いたことを感謝します。

私の日本での勉強を手配して頂いた中国教育委員会そして、修士課程と博士課程の奨学金を提供して頂いた日本文部省に深く感謝します。

最後に、大学の生活及び勉強を支えてくれた妻に感謝します。

参考文献

- [1] G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," Proc. Int. Conf. IEEE Computer Vision and Pattern Recognition, pp. 70-77, June 1985.
- [2] J. Aloimonos and A. Bandyopadhyay, "Active vision," Proc. ICCV'87, pp. 35-55.
- [3] D. H. Ballard, "Reference frame for animate vision," Proc. IJCAI-89, pp. 1635-1641, 1989.
- [4] Q. Chen and S. Tsuji, "A generalized shape and motion from image sequences," Proc. First Korea-Japan Joint Conf. Computer Vision, pp. 343-349, October 10-11, 1991.
- [5] A. R. Bruss and B. K. P. Horn, "Passive navigation," Computer Vision, Graphics and Image Processing, vol. 21, no. 1, pp. 3-20, Jan. 1983.
- [6] O. D. Faugeras, F. Lustman, and G. Toscani, "Motion and structure from motion from point and line matches," Proc. 1st Int. Conf. Computer Vision, pp. 25-34, June 1987.
- [7] Fredrik Bergholm, "Decomposition theory and transformations of visual directions," Proc. ICCV'90, pp 85-90.
- [8] T. C. Chou and K. Kanatani, "Recovering 3-D rigid motions without correspondence," Proc. ICCV'87, pp. 534-538.
- [9] G.H. Golub and C. Reinsch, "Singular value decomposition and

- least squares solutions," Handbook for Automatic Computation vol. 2, Linear Algebra, chapter I/10, pp 134-151, Springer Verlag, 1971.
- [10] H. H. Hagel, "Image sequences-ten(octal) years-from phenomenology towards a theoretical foundation," Proc. ICPR'86, pp. 1174-1185.
- [11] B. K. P. Horn, Robot Vision, Cambridge, MA: MIT Press, 1986.
- [12] B. K. P. Horn and E. J. Weldon, "Computationally efficient methods for recovering translational motion," Proc. ICCV'87, pp. 2-11.
- [13] T. S. Huang and R. Y. Tsai, "Image sequence analysis: motion estimation," in Image Sequence Processing and Dynamic Scene Analysis, T. S. Huang, Ed. New York, NY: Springer-Verlag, 1981.
- [14] D. T. Lawton, "Processing translational motion sequences," Computer Vision, Graphics and Image Processing, vol. 22, pp. 116-144, 1983.
- [15] Y. C. Liu and T. S. Huang, "Estimation of rigid body motion using straight line correspondences," Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis, pp. 47-52, May 1986.
- [16] Y. C. Liu and T. S. Huang, "Estimation of rigid body mo-

- tion using straight line correspondences: further results," Proc. ICPR-86, pp. 306-309, Oct. 1986.
- [17] H. C. Longuet-Higgins, "A computer algorithm for reconstruction a scene from two projections," Nature, vol. 293, pp. 133-135, Sep. 1981.
- [18] H. C. Longuet-Higgins, "The reconstruction of a scene from two projections-configurations that defeat the 8-point algorithm," Proc. 1st Conf. on Artificial Intelligence Application, pp. 395-397, Dec. 5-7, 1984.
- [19] H. H. Nagel, "Representation of moving rigid objects based on visual observations," Computer, pp. 29-39, Aug. 1981.
- [20] S. Negahdaripour and B. K. P. Horn, "Determining 3-D motion of planar objects from image brightness measurements," Proc. IJCAI-85, pp. 898-901.
- [21] K. Prazdny, "Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer," Computer Graphics and Image Processing, vol. 17, pp. 238-248, 1981.
- [22] S. Negahdaripour and B. K. P. Horn, "Direct passive navigation," IEEE Trans. Pattern Anal. Machine Intell., vol. PAML9, no. 1, pp. 168-176, Jan. 1987.
- [23] J. H. Rieger and D. T. Lawton "Determining the instantaneous

- axis of translation from optic flow generated by arbitrary sensor motion," Proc. ACM Interdisc. Workshop Motion, pp. 33-44, 1983.
- [24] J. W. Roach and J. K. Aggarwal, "Computer tracking of objects moving in space," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-1, no. 2, pp. 127-135, Apr. 1979.
- [25] J. W. Roach and J. K. Aggarwal, "Determining the movement of objects from a sequence of images," IEEE Trans. Pattern Anal. Machine Intell. vol. PAMI-2, no. 6, pp. 554-562, Nov. 1980.
- [26] M. Subbarao, "Solution and uniqueness of image flow equations for rigid curved surfaces in motion," Proc. ICCV-87, pp. 687-692.
- [27] M. E. Spetsakis and J. Y. Aloimonos, "Optimal motion estimation," Proc. the IEEE Workshop on Visual Motion, pp 229-237, March 1989.
- [28] C. Tomasi, T. Kanade, "Shape and motion without depth," Proc. ICCV'90, pp 91-94.
- [29] R. Y. Tsai and T. S. Huang, "Estimating 3-D motion parameters of a rigid planar patch, I," IEEE Trans. Acoust. Speech Signal Processing, vol. ASSP-29, no. 6, pp. 1147-1152, Dec. 1981.

- [30] R. Y. Tsai, T. S. Huang, and W. L. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition," IEEE Trans. Acoust Speech Signal Processing, vol. ASSP-30, pp. 525-534, Aug. 1982.
- [31] R. Y. Tsai, and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces," PAMI, 6(1) 13-26, 1984.
- [32] S. Ullman, "The interpretation of visual motion," Cambridge, MA: MIT Press, 1979.
- [33] S. Ullman, "The interpretation of structure from motion," Proc. R. Soc. London, B203, pp. 405-426, 1979.
- [34] S. Ullman, "Maximizing rigidity: the incremental recovery of 3D structure from rigid and non-rigid motion," Perception, vol. 13, pp. 255-274, 1984.
- [35] A. M. Waxman, B. Kamgar-Parisi, and M. Subbarao, "Closed the form solutions to image flow equations," Proc. of the 1st Conf. on Artificial Intelligence, pp. 12-23, 1984.
- [36] A. M. Waxman and K. Wohn, "Image flow theory: a framework for 3-D inference from time-varying imagery," in Advances in Computer Vision, C. Brown, Ed. Hillsdale, Nj: Erlbaum Publishers, 1987.
- [37] T. D. Williams, "Depth from camera motion in a real world

- scene," IEEE Trans. Pattern Anal, Machine Intell., vol. PAMI-2, no. 6, pp. 511-516, Nov. 1980.
- [38] B. L. Yen and T. S. Huang, "Determining 3-D motion and structure of a rigid body using straight line correspondences," in Image Sequence Processing and Dynamic Scene Analysis, T. S. Huang, Ed. New York, NY: Springer Verlag, 1983.
- [39] B. L. Yen and T. S. Huang, "Determining 3-D motion/structure of a rigid body over 3 frames using straight line correspondence," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 267-272, June 19-23, 1983.
- [40] J. Y. Zheng, Q. Chen and S. Tsuji, "Active camera guided manipulation," Proc. IEEE Int. Conf. Robotics and Automation 1991, pp. 632-638.
- [41] J. Y. Zheng, Q. Chen, F. Kishino and S. Tsuji, "Active camera controlling for manipulation," Proc. Int. Conf. CVPR 1991, pp. 413-418.
- [42] X. Zhuang, T. S. Huang, and R. M. Haralick, "Two-view motion analysis: a unified algorithm," J. Opt. Soc. Amer., vol. 3. no. 9, pp. 1492-1500, Sept. 1986.
- [43] X. Zhuang and R. M. Haralick, "Two view motion analysis-theory and algorithm," Proc. ICASSP, Mar. 1985.
- [44] X. Zhuang and R. M. Haralick, "Two view motion analysis,"

- Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 686-690. June. 1985.
- [45] T. J. Brodia and R. Chellappa, "Estimation of object motion parameters from noisy images," IEEE Trans. Pattern Anal. Machine Intell. vol. PAMI-8, no. 1, pp. 90-99, Jan. 1986.
- [46] M. Subbarao, "Interpretation of image motion fields: rigid curved surfaces in motion," Tech. Rep. CAR-TR-199, Center for Automation Research, Univ. of Maryland, College Park, MD, April, 1986.
- [47] H. Ishiguro, P. Stelmazyk and S. Tsuji, "Acquiring 3-D structure by controlling visual attention of a mobile robot," Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 755-760, 1990.

研究業績

- [1] Q. Chen, M. Asada and S. Tsuji, "A new 2-D representation system for mobile robot," Proc. 9th International Conference on Pattern Recognition, 1988, pp. 604-606.
- [2] Q. chen, M. Asada and S. Tsuji, "A new polygon-based world representation system for mobile robots," 第6回日本ロボット学会学術講演会論文集, 1988年10月20-22日.
- [3] J. Y. Zheng, Q. Chen and S. Tsuji, "Active camera guided manipulation," Proc. IEEE Int. Conf. Robotics and Automation 1991, pp. 632-638.
- [4] J. Y. Zheng, Q. Chen, F. Kishino and S. Tsuji, "Active camera controlling for manipulation," Proc. Int. Conf. CVPR 1991, pp. 413-418.
- [5] Q. Chen and S. Tsuji, "A generalized shape and motion from image sequences," Proc. First Korea-Japan Joint Conf. on Computer Vision, pp. 343-349, October 10-11, 1991.
- [6] 陳謙, 辻三郎, "更新効率を重視した屋内移動ロボットの経路計画の為の新しい地図モデル," 日本ロボット学会学会誌, 1992. (採録決定)
- [7] 陳謙, 辻三郎, "画像列から環境構造及びカメラ運動の決定の新しい手法," 電子通信情報学会論文誌, 1992. (投稿中)

- [8] Q. Chen and S. Tsuji, "A new approach to the computation of structure and motion from image sequence," Proc. CVPR-92. (投稿中)
- [9] Q. Chen and S. Tsuji, "A hierarchical method that solves the shape and motion from image sequences problem," Proc. IEEE International Conference on Robotics and Automation, 1992. (投稿中).

