

移動ロボットに搭載したカメラの視点移動による 物体の3次元形状モデリング

山崎 公俊 (筑波大) 友納 正裕 (筑波大) 坪内 孝司 (筑波大) 油田 信一 (筑波大)

3-D Object Modeling by a Camera Equipped on a Mobile Robot

*Kimitoshi Yamazaki, Masahiro Tomono, Takashi Tsubouchi and Shin'ichi Yuta
(University of Tsukuba)

Abstract— This paper describes 3-D shape modeling performed by a mobile robot. We propose a method in which the robot in the real environment acquires the accurate shape and constructs a 3-D model of an object using Computer Vision techniques under the condition that position and shape of the object is unknown. 3-D modeling is performed from image streams which are captured by a camera placed on the robot. The results show the effectiveness of our method.

Key Words: 3-D Shape Modeling, Vision, Mobile Robot, Single Camera

1. はじめに

オフィスや一般家庭などの実環境を動作する移動ロボットには、周囲の環境を観察することで、行動を決定し実行することが求められる。その際、ロボットに必要な機能のひとつは、周囲の情報を行動に役立つ形態で保持することである。近年、移動ロボットに搭載したセンサより得た情報をもとに、未知の環境において地図を生成するマップ構築の研究がおこなわれている [6][11]。これらの多くはナビゲーションを目的としており、環境の大域的な情報が取得できる一方、環境内の特定の物体を操作できるような詳細な情報は得られない。

筆者らは、実環境に存在する物体に対し、ひろいあげのような物体操作を自律移動ロボットで実現することに興味がある。このためには、個々の操作対象物体についての詳細な形状情報を持つことが必要であるが、実環境には多くの物体が存在しており、それら全ての形状モデルをロボットにあらかじめ持たせておくことは困難である。本研究ではロボットに、対象物の位置や形状が未知の状態から、実環境中の物体の詳細な形状モデルを生成することを目指す。

本論文は、実環境で動作する自律移動ロボットによる、把持・操作が可能な小型物体の3次元形状モデリングについて述べる。単眼のビジョンセンサ(パン・チルトカメラ)を搭載した車輪型移動ロボットにより対象物に関する画像列を取得し、因子分解法、モーションステレオ、Space Carvingなどのコンピュータビジョン手法を用いて、対象物の形状モデリングをおこなう手法を提案する。

本手法は、以下のような特長を持つ。

- ロボットが自律的に対象物を探索し決定する。
- ロボットが走行しながら形状モデルを生成する。
- 対象物には人工的なマーカを持たせない。
- ロボットの動作モデルなしに、画像列のみから処理をおこなう。

以下では、2章で関連研究について述べ、3章で本研究の方針と問題設定について述べる。4章では3次元形状モデリングの概要、5章では提案する密な形状復元手法の詳細、6章で提案手法を実装した実験結果を示す。7章をまとめとする。

2. 関連研究

複数の画像データから物体の3次元形状を復元する研究は、コンピュータビジョンの分野で盛んにおこなわれている。

因子分解法 [1][8] や、エピポーラ幾何を用いた手法 [15]、拡張カルマンフィルタを用いた手法 [2] は、複数の画像デー



Fig.1 車輪型移動ロボット

タから、物体の疎な3次元形状と、カメラの位置・姿勢を得ることができ、画像間でテクスチャの対応が付く場合には有用な手法である。また、物体の密な形状を復元する手法として、相関ステレオ法 [7] や、Space Carving [5] などがある。これらは、カメラの位置・姿勢が既知である場合に利用できる。

種々の手法を統合して、カメラの動きデータ無しに画像列から物体の密な形状を復元する研究もおこなわれている。Satoら [10] は、あらかじめ指定したマーカと自動で追跡した特徴点からカメラ姿勢を求め、それらを元にマルチベースラインステレオ法、Voxel空間への統合をおこない、屋外環境の密な3次元形状モデルを生成した。Sainzら [9] は、透視投影モデルに基づく因子分解法により求めたカメラの位置・姿勢を用いて、Space Carvingをおこない、密な形状を得た。

対象物や背景に条件を課し、3次元形状を精密に復元しているこれらの研究に対し、本研究では、実環境を背景に、物体には人工的なマーカが無い状態から、移動ロボットが自動で3次元形状モデリングをおこなう。

3. 研究の方針と問題設定

3-1 問題設定

本研究では、車輪型移動ロボット (Fig.1) が床面上を移動することにより、複数の視点から画像列を取得し、それらを用いて形状モデリングをおこなう。

対象物に関しては、以下の仮定をおく。

1. 対象物はロボットが移動する床面上に存在する

2. 対象物の形状は曲面を含んでも良いが、テクスチャ（模様）がある。

特に 2. の仮定は、画像列から形状情報を得る際、画像中に映り込んだ対象物のテクスチャを利用するためである。

3.2 課題とアプローチ

実環境を移動するロボットに自律的に形状モデリングをおこなわせる際、以下の難しさがある。

対象物の条件

対象物には、ロボットが他の物体と区別が付けられるような特別な条件や、形状取得を補助するようなマーカなどを持たせない。すなわちロボットは、形状モデリングをおこなう対象物を、実環境中に存在する物体の中から特定する必要がある。また、形状モデリングの際にも、カメラから得た画像中から、モデリングに必要な情報を随時抽出しなければならない。

カメラの位置・姿勢の誤差

移動ロボットに搭載したカメラの位置・姿勢を直接的に知る手段は、オドメトリ（車輪の回転数から測位するロボットの自己位置）とパン・チルトの角度情報である。しかし、これらから得るカメラの位置・姿勢は、画像取得とオドメトリ取得の同期のずれや、車体のスリップ、カメラの角度の誤差などの要因により、正確ではない。

すなわち、ロボットは最初実環境を広い範囲で探索し、環境中に存在する複数の物体から、形状モデリングの対象物を決定する必要がある。そして対象物の詳細な形状を得るためには、その物体の周囲を局所的に探索し、カメラの位置・姿勢の誤差にロバストな形状モデリングをおこなう。以上より、本研究では処理を以下の 2 段階に分ける。

1. 対象物を移動しながら発見する。
(対象物の発見)
2. 発見した物体の形状モデルを生成する。
(対象物の形状モデリング)

対象物の発見処理では、ロボットは環境中を移動しながら、形状モデリングの対象物を自動で決定する。すなわち、ロボットは床面上の物体の有無を探索しながら走行し、周囲の物体の中から決定した対象物について、その正確な位置と、大まかな形状を求める。この詳細は文献 [12] に譲る。対象物の形状モデリングでは、ロボットは発見した対象物に近づき、その周囲を走行することで画像列を取得する。これらの画像列から、対象物の密な形状を獲得する。本論文では形状モデリングの詳細について述べる。

カメラの位置・姿勢の推定は、オドメトリから得る測位情報のみでは不十分であるため、その誤差を修正する方法、もしくはオドメトリ情報に頼らずカメラの位置・姿勢を求める手法をとる。

4. 対象物の形状モデリング

本研究での形状モデリングの要件を以下に示す。

1. 移動ロボットが走行中に可能な処理である
2. 形状の復元結果によりロボットの走行経路が変更可能、すなわち逐次処理である

ロボットが動作する環境において、対象物を発見した際、判明している対象物の 3 次元形状は一部でしかなく、形状モデリングにより対象物の正確な全体像を得ることが必要である。本研究は、対象物に対しひろいあげなどの物体操作を想定しているため、形状モデリングには高い復元精度が要求される。

本章では、以前筆者らが提案した、画像列からカメラの位置・姿勢と対象物の疎な形状を逐次的に得る手法 [14] を改良し、密な形状復元を加えたアルゴリズムについて述べる。コンピュータビジョンの既存の手法については、付録に示す。

4.1 画像列からの形状モデリング手法の概要

4.1.1 特徴点の抽出・追跡

対象とする物体はテクスチャを有するため、テクスチャから特徴点を抽出し、その特徴点が次に取得した画像上のどこに移動したかを調べ、追跡する。本研究では、特徴点を抽出・追跡するために KLT-Tracker[4] を用いる。これにより、物体の特徴点の画像間での推移を記録する。

4.1.2 カメラ運動と物体形状の復元

車体やカメラの動作データから得るカメラの位置・姿勢には誤差が含まれることから、カメラの位置・姿勢と対象物の形状を画像列を用いて求める。具体的には、KLT-Tracker により得た特徴点の推移情報に対し、コンピュータビジョン分野の種々の手法を、それぞれの特性と画像列の状況に応じて使い分けて利用することで、カメラの位置・姿勢と特徴点の 3 次元位置を精度良く求める。

本研究では、初期フェーズにおいて、カメラ位置・姿勢と特徴点の 3 次元位置を同時に算出可能な因子分解法を利用し、逐次フェーズでは非線形最小化 [3] を用いてカメラの位置・姿勢を、モーションステレオ [3] を用いて特徴点の 3 次元位置を求めている。それぞれのフェーズについては次節で述べる。

4.1.3 対象物の密な形状復元

画像を取得した際のカメラの位置・姿勢が既知であれば、相関ステレオなどの手法により、対象物の密な 3 次元形状を得ることができる。本論文では、カメラの位置・姿勢に加え、画像間での特徴点の推移情報も既知であることを利用した、密な形状復元手法を提案する。この詳細は 5 章で述べる。

4.2 形状モデリングアルゴリズム

本節では、形状モデリングアルゴリズム (Fig.2) の詳細を説明する。

前提条件

対象物の発見処理により、物体とロボットの相対位置と、物体の大まかな形状が判っている。ロボットは、まず発見された物体に直線的に近づき、その後は対象物とカメラとの距離を所定の範囲内に保つように物体の周囲を走行する。

初期フェーズ

逐次フェーズのための初期値を生成するフェーズである。

対象物の発見処理で求められた特徴点の 3 次元位置や、オドメトリから得るカメラの位置・姿勢には、誤差が多く含まれる。これらを正確に求め直すため、複数枚の画像中の複数個の特徴点を用いて、因子分解法をおこなう。その後、因子分解法の誤差を非線形最小化によって補正し、正確なカメラの位置・姿勢と、複数の特徴点の 3 次元位置をまとめて求める。

発見処理の後、対象物に接近しながら走行する間に取得した画像列では、映り込む物体の姿勢変化が少なく、オクルージョンが起こりにくい。そのため、多くの特徴点が複数枚の画像で観測し続けられる。この理由から、初期フェーズでは因子分解法の利用が有効である。複数枚の画像にわたって追跡が続いた複数の特徴点の推移情報を用いて、因子分解法によりカメラの位置・姿勢と特徴点の 3 次元位置を同時に算出する。その結果を初期値とし、非線形最小化を用いて補正処理を加える。因子分解法の結果は線形近似による誤差を含むが、非線形最小化の初期値としては、十分に良い近似値となる。

逐次フェーズ

対象物の周囲から得た画像から、その全体形状を獲得するフェーズである。このフェーズでは、新しく画像を取得するたびに以下の処理をおこなう。

- 3 次元位置が既知で、画像間で追跡が続いている特徴点から、カメラの位置・姿勢を非線形最小化により推定

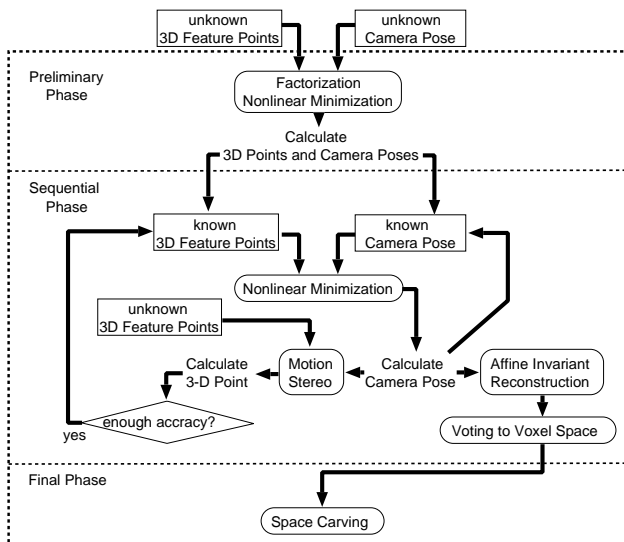


Fig.2 形状モデリングアルゴリズム

- 求めたカメラの位置・姿勢から、新たに抽出された特徴点の3次元位置をモーションステレオにより算出
- 求めたカメラの位置・姿勢から、アフィン不変量を利用した密な形状復元をおこない、Voxel空間に投票

対象物の周囲の画像を取得する段階に入ると、物体の見え方の変化に伴い、特徴点は頻繁に入れ替わる。このため、特徴点の3次元位置算出を1点ごとにおこなえるモーションステレオが有効である。カメラの位置・姿勢は、画像中に存在し3次元位置が既知の特徴点から、非線形最小化により推定する。

逐次フェーズでは、取得した画像で特徴点を追跡した後、3次元位置が既知の特徴点を用いてカメラの位置・姿勢を算出する。具体的には、現在のカメラの位置・姿勢のパラメータを変数とし、一つ前のカメラ位置・姿勢を初期値とした非線形最小化をおこなう。

このとき、Trackerが追跡ミスを起こした特徴点をもとに非線形最小化をおこなうとカメラの位置・姿勢が正確に求まらない。この対策としてRANSACの枠組みを用いる[13]。具体的には、3次元位置が既知である画像特徴点の6割程度をランダムに選び出し、非線形最小化によりカメラの位置・姿勢を推定することを数百回程度繰り返す。最小化の収束値が一番小さくなるカメラの位置・姿勢を推定値とする。

3次元位置が未知の特徴点は、モーションステレオにより復元する。このときの復元の正確さは、求めた3次元位置を現在の画像上に逆投影し、その投影結果とTrackerによる特徴点の追跡結果との差分を観察することで判定する。差分が所定の閾値以下であれば、その特徴点を既知とし、以降のカメラの位置・姿勢の推定に利用する。一方で、複数枚の画像毎にアフィン不変量を利用した密な形状復元をおこない、結果をVoxel空間に投票する(5章参照)。

このフェーズでは、それぞれの処理が高速であり、画像を取得するたびに対象物の形状を逐次的に求めることができる。そのため、処理の進行と共に判明する対象物の形状に合わせて、形状モデリングをより詳細におこなえるようなカメラの移動経路を、同時に生成することが可能である。

最終フェーズ

逐次フェーズまでに取得した対象物の全体形状を補正するフェーズである。Voxel空間の投票結果から、所定の閾値を越えるVoxelのみを残す。そのVoxel集団をSpace Carving[8]により削ることで、最終的な形状を得る。

逐次フェーズにより得たVoxelによる対象物の形状には、アフィン変換の利用による誤差が含まれる。最終フェーズでは、特徴点の推移情報を用いて高速に求めた3次元形状を、

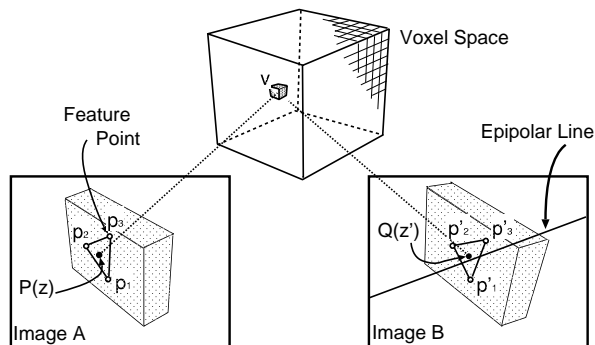


Fig.3 アフィン不変量の利用

Photometricな情報に基づくSpace Carvingによって補正する。詳細は、5章に述べる。

5. 密な形状獲得

本章では、カメラの位置・姿勢と、画像間で特徴点の対応が求まっている条件をもとで、提案する密な形状復元手法について述べる。

5.1 アフィン不変量を利用した密な形状復元

カメラの位置・姿勢が既知であれば、2枚の画像間で各画素の対応を取ることで、密な3次元形状を復元できる。本研究では、カメラの位置・姿勢に加え特徴点の対応も既知であることから、密な形状復元を高速におこなう手法を提案する。

ある画像内において指定した画素に対し、その近傍の特徴点3点から、その画素を内包するような3角形を構成する。指定した画素と3点の特徴点から、以下の式によりアフィン不変量を算出する。不変量とは、あるモデルの変形前と変形後において、その間の変形の変換に対し不変な表現を意味する。

$$z = \alpha(p_2 - p_1) + \beta(p_3 - p_1) + p_1 \quad (1)$$

ここで、 z は指定した画素、 $p_n (n=1,2,3)$ は、画像内の特徴点であり、 α, β はアフィン不変量を表す。(1)式から、 α, β が算出できる。一方、画像A中の z に対する画像B中の推定座標 z' は、以下の式で表される。

$$z' = \alpha(p'_2 - p'_1) + \beta(p'_3 - p'_1) + p'_1 \quad (2)$$

$p'_n (n=1,2,3)$ は、画像A中の特徴点に対応付けられた画像B上の特徴点を表す。(2)式では、右辺のパラメータは全て既知となるため、 z' が一意に求まる。

この枠組みを利用することで、現在の画素(Fig.3, P)が別の画像上のどこに移動したかを、算出したアフィン不変量と3角形を構成する近傍の特徴点の位置から求めることができる(Fig.3, Q)。ただし、本手法では2次元のアフィン不変量を用いているため、画素の推定位置には近似による誤差が生じる。そこで以下の条件から画素の推定位置を評価し、その結果をもとに復元する画素を選ぶことで、誤差の大きな点が復元されることを制限する。

- 推定した画素の座標が、もとの画素の座標によるエピ極線から所定の閾値以内の距離にある。
- 推定した画素の輝度値が、実際の画素の輝度値に近い。

以上より、2枚の画像間での画素の対応を付け、モーションステレオと同様の計算をおこなうことで、3次元位置を求める。

この手法は、特徴点の対応が付いていれば、相関により画素の対応付けをする必要がなく、高速に密な3次元復元が可能である。

5.2 Voxel空間への統合

前節の手法により密な復元をおこなった結果を、Voxel空間に統合する。アフィン不変量を利用した復元手法は、2次

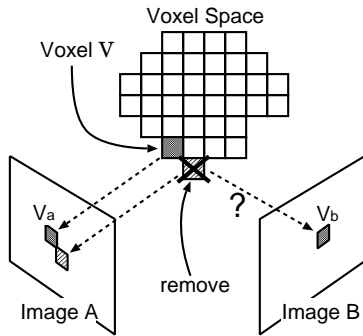


Fig.4 Space Carving

元のアフィン変換により近似をおこなっているため、奥行き
の大きなシーンや、3次元形状に細かい凹凸のある部分では、
復元誤差が大きくなる。一方で、本研究では多数枚の画像を用
いて対象物の形状モデリングをおこなうため、2枚の画像毎
に密な形状復元を続けると、冗長なデータが非常に多くなる。

そこで、物体の周囲の画像から2枚の画像毎に復元をおこ
ない、その結果を Voxel 空間に統合する。具体的には、対象
物の大きさをもとに定義した Voxel 空間において、2枚の画
像から復元した画素の3次元位置を含んでいる Voxel に対し、
投票する。このとき、画素の3次元位置にはアフィン変換の
誤差が含まれていることを考慮し、近傍の Voxel にも投票す
る。他の画像間でも同様の処理をおこない、最終的な投票値
が所定の閾値を越えた Voxel のみを残す。得られる復元結果
は、形状に厚みを持った Voxel の集団として表される。

5.3 Space Carving による形状の補正

上記の処理で得られた Voxel 集団に対し、Space Carving[8]
をおこなう。Space Carving とは本来、カメラ位置が既知であ
る任意の視点から得た複数枚の色画像から、物体の3次元形
状を Voxel 集団として復元する手法である。

この手法を利用する際、Carving をおこなうための初期形状
をどのように求めるかが課題となる。通常は、対象物と区別
しやすい背景のもとで画像を取得し、輪郭に基づく制約(シル
エット制約)を利用して初期形状を生成することが多いが、
本研究では、ロボットが移動するのは実環境であり、取得す
る画像から対象物の輪郭のみを抽出することは困難である。
しかし、前節の方法で対象物の大まかな形状が Voxel 集団と
して得られているため、これを初期形状とできる。

この初期形状を、Photometric な情報に従い削る。具体的
には、Fig.4 に示すように、ある Voxel V が画像 A 上に投影さ
れた際、その投影像 v_a に含まれる画素の輝度値を記録してお
く。そして別の画像 B 上にその Voxel V が投影された際、そ
の投影像 v_b の画像領域に含まれる輝度値を調べ、画像 A の
 v_a の輝度値と異なれば、その Voxel は消去する。これを複数
枚の画像間で繰り返し、最終的な対象物の形状を求める。

Space Carving は、幾何学的な近似を含まず、画像の Pho
tometric な情報により処理をおこなうため、アフィン不変量
を利用した密な形状復元の誤差を補正できる。

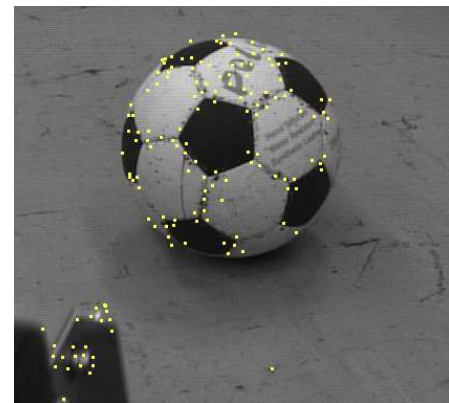
6. 実験

2輪駆動型の移動ロボット「山彦」に、Canon VC-C4 パン
チルトカメラを搭載したシステムを用いて実験をおこなった。
対象物 (Fig.5) は、直径 220mm の球体である。画像取得は、
HITACHI 製の画像処理モジュール「IP7500」を用いた。画像
の解像度は 512×440 pixel であった。これらの画像列に対す
る処理を、Pentium4, CPU2.8 GHz の計算機を用いてオフラ
インでおこなった。

床面上に設置した対象物に対し物体から 1500mm 離れた位
置にロボットを設置し、ロボットには、物体に向かって走行
させた後、その周囲に半径 800mm の円経路を与えた。画像



Original Image



Feature Tracking

Fig.5 原画像と特徴点抽出

はロボットが 20mm 進む毎に取得した。ロボットの走行経路
と画像取得の間隔については、画像間で視差が大きな場合や
スケールの変化が大きい場合に、KLT-Tracker が追跡ミス
を起こしやすいことから、視差が十分小さくなるよう経験的
に決定した。カメラの位置・姿勢は、全ての画像において推
定している。形状復元については、因子分解法を 20mm 毎
に取得した複数枚の画像を用いておこない、一方で、モーシ
ョンステレオによる3次元復元には視差が必要なため、2枚
の画像取得間隔が 60mm 以上離れている場合におこなった。

取得した対象物の全周囲画像群を用いて、初期フェーズ、
逐次フェーズ合わせて 1000 点程度の特徴点の3次元位置と、
200ヶ所のカメラの位置・姿勢を復元した。最終フェーズに
おいて、3次元形状を一辺 2mm の Voxel により求めた。Space
Carving により最終的に残った Voxel の数は、約 60000 個
であった。処理時間については、逐次フェーズの各画像にお
いて、カメラの位置・姿勢の算出に 0.7 sec、アフィン不変量
を利用した復元に 0.25 sec であり、1枚の画像につきほぼ 1
sec 以内で処理可能である。最終フェーズにおいては、Space
Carving に約 180sec を要した。また、物体の直径を実測値と
比較した際、モデリング結果の誤差は 10mm 程度で納まる
ことを確認した。形状モデリングの結果の例を Fig.6、Fig.7
に示す。Fig.6、(1) は、逐次フェーズにより得られた特徴
点の3次元位置であり、物体の疎な形状が復元されている。
Fig.6、(2) は、逐次フェーズにおける2枚の画像間での密
な復元の結果である。球体の一部の形状が得られており、
アフィン変換による誤差が含まれていることが判る。これら
の結果を Voxel 空間に統合し、Space Carving をおこな
った結果、最終的に Fig.7 に示すように良好な復元結果
が得られた。

7. まとめ

本稿では、移動ロボットに搭載するための、対象物の形状
モデリング手法を提案した。また、実際に移動ロボットを走

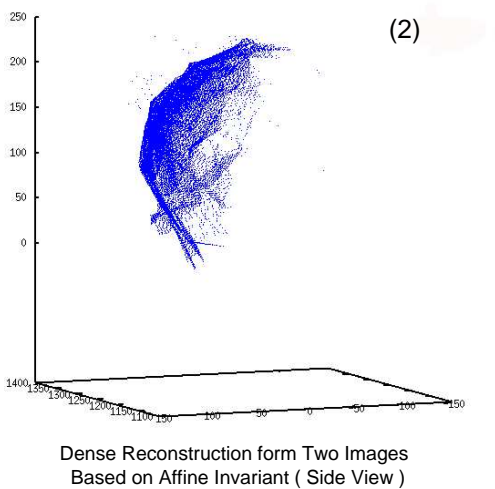
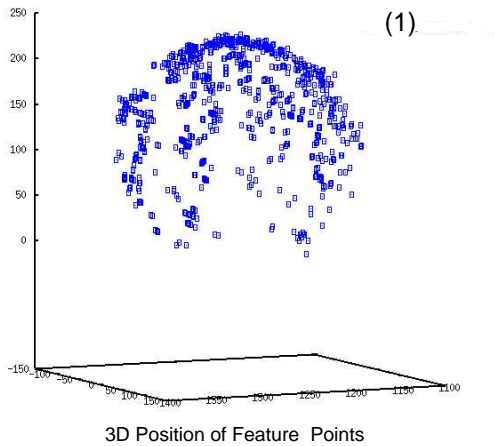


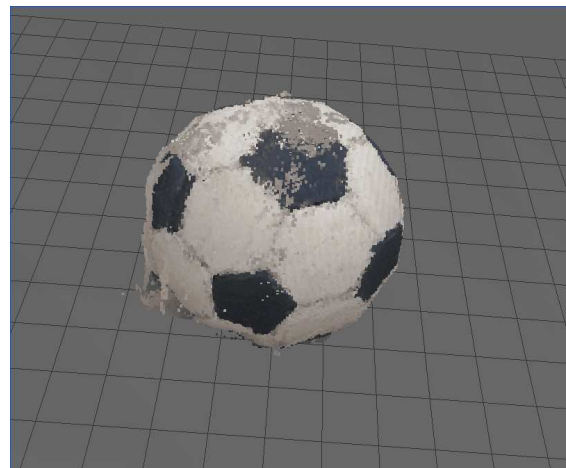
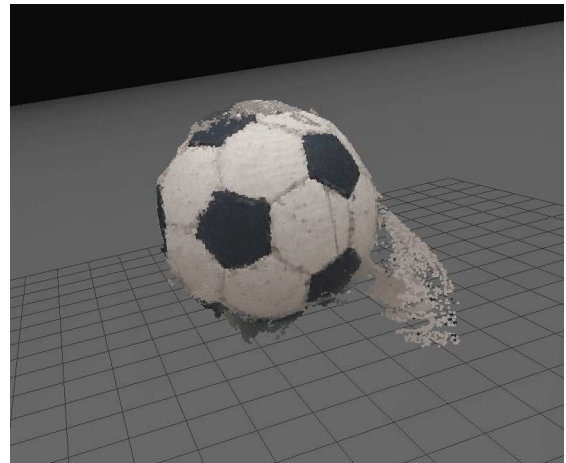
Fig.6 形状モデリングアルゴリズムの中間結果

行させながら取得した画像列を用いて、実画像による実験から、数分のオーダーで3次元形状モデリングをおこない、提案手法の有効性を示した。

今後の課題として、形状モデリングに最適となるような、ロボットの視点移動計画がある。また、処理を高速化し、移動ロボットの自律的な形状モデリングの実現を目指す。

参考文献

- 1) C. Tomasi, T. Kanade: "Shape and Motion from Image Streams under Orthography: a Factorization Method", IJCV,9(2), pp.137-154, 1992.
- 2) 出口: 「ロボットビジョンの基礎」, コロナ社, 2000.
- 3) 徐: 「写真から作る3次元CG-イメージ・ベースド・モデリング&レンダリング-7章,8章」, 近代科学社, 2002.
- 4) J. Shi, C. Tomasi: "Good Features to Track", CVPR, 1994.
- 5) Kutulakos K. and Seitz S.: "A Theory of Shape by Space Carving", IJCV, 38(3), pp.198-218, 2000.
- 6) L. Iocchi, D. Mastrantuono, D. Nardi: "Visually Realistic Mapping of a Planar Environment with Stereo" ISER 2000.
- 7) M. Okutomi. and T. Kanade: "A Multiple-Baseline Stereo", PAMI, VOL15,NO.4, pp.353-363, 1993.
- 8) Poelman C.J and Kanade T.: "A Paraperspective Factorization Method Shape and Motion Recovery" tech. report CMU-CS-93-219, Computer Science Department, Carnegie Mellon University, 1993
- 9) Sainz M., Bagherzadeh N. and Susin A.: "Carving 3D Models from Uncalibrated Views.", CGIM 2002, pp.144-149., 2002.



3D Modeling Result

Fig.7 3次元形状モデリングの結果

- 10) T. Sato, M. Kanbara, N. Yokoya, H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Handhelds-baseline Stereo using a Hand-held Video Camera", IJCV, Vol 47, No.1-3, 2002.
- 11) S. Thrun, W. Burgard, and D. Fox: "A Probabilistic Approach to Concurrent Mapping and Localization for a Mobile Robot", Machine Learning and Autonomous Robots, 31/5, 1-25, 1998.
- 12) 山崎, 友納, 坪内, 油田: 「移動マニピュレータによる物体ひろいあげに関する研究」, 第8回ロボティクス・シンポジウム, 2003.
- 13) 山崎, 友納, 坪内, 油田: 「移動ロボットによる物体ひろいあげのための対象物の3次元形状獲得」, SI2003, 2003.
- 14) 山崎, 友納, 坪内, 油田: 「移動マニピュレータによる物体ひろいあげのための対象物の逐次的な形状復元」, 第21回日本ロボット学会学術講演会, 2003.
- 15) Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong.: "A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry", Artificial Intelligence Journal, Vol.78, pp.87-119, 1995.

付録

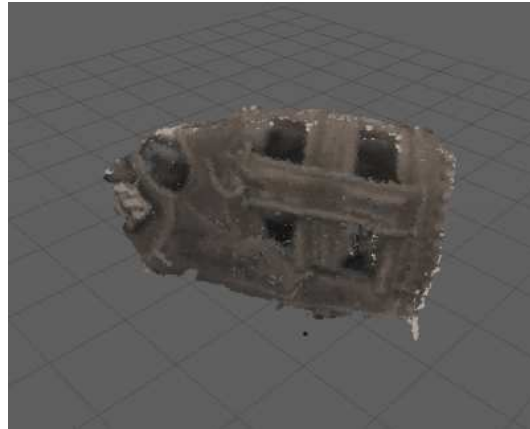
モーションステレオ

2枚の画像間に対応している特徴点があり、画像取得時のカメラの位置・姿勢がそれぞれ既知であるとき、モーションステレオが可能となる。本研究では、以下の関数を線形最小化することにより、特徴点の3次元位置を求める。

$$C = \|\mathbf{X} - s_1 \tilde{\mathbf{m}}_1\|^2 + \|\mathbf{X} - s_2 \mathbf{R}\tilde{\mathbf{m}}_2 + \mathbf{T}\|^2 \quad (3)$$



Original Image



3D Modeling Result

Fig.8 他のモデリング例

$\tilde{\mathbf{m}}_1, \tilde{\mathbf{m}}_2$ は、2枚の画像における対応点の投影座標の拡張ベクトルである。 $\mathbf{X} = (X, Y, Z)$ は特徴点の3次元位置、 \mathbf{R} は2枚の画像間でのカメラの相対姿勢、 \mathbf{T} は相対位置である。

この手法は特徴点1点毎に処理が可能であり、高速な処理が期待できる。ただし、カメラの位置・姿勢に誤差が含まれると、それが復元結果に大きく影響する。そのため、形状モデリングの初期フェーズでは利用しない。

因子分解法

因子分解法とは、複数の画像から得られた複数の特徴点を並べた行列を作成し、ある条件の下で因子分解をおこなうことにより、カメラの姿勢と、特徴点の3次元位置を同時に求める方法である。以下に概要を述べる。F枚の画像があり、全ての画像間において対応が付いている特徴点がP個あるとする。特徴点pのfフレーム目の画像上での位置を (u_{fp}, v_{fp})

としたとき、以下の行列 \mathbf{W} を定義する。

$$\mathbf{W} = \begin{bmatrix} u_{11} & \cdots & u_{1P} \\ \vdots & & \vdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ \vdots & & \vdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} \quad (4)$$

この $2F \times P$ の行列を、計測行列と呼ぶ。ここから、 \mathbf{W} の各行毎に平均を求め、それを \mathbf{W} から差し引いた行列 \mathbf{W}' をつくる。そして、カメラ姿勢を、ワールド座標系における3つの正規直交規定 $\mathbf{i}_f = (i_{fx}, i_{fy}, i_{fz})$ 、 $\mathbf{j}_f = (j_{fx}, j_{fy}, j_{fz})$ 、 $\mathbf{k}_f = (k_{fx}, k_{fy}, k_{fz})$ とし、空間上の点の座標を $\mathbf{s}_p = (s_{px}, s_{py}, s_{pz})$ としたとき、(1) $\mathbf{i}_f, \mathbf{j}_f$ は直交する、(2) $\mathbf{i}_f, \mathbf{j}_f$ の長さは1である、等の拘束条件を用いることで、特徴点列から作成した行列 \mathbf{W}' との間に次に式が成り立つ。

$$\mathbf{W}' = \mathbf{M}\mathbf{S} \quad (5)$$

$2F \times 3$ の行列 \mathbf{M} はカメラの運動を表しており、 $3 \times P$ の行列 \mathbf{S} は特徴点の3次元位置を表している。

この手法は、カメラ位置を既知とせず、画像情報のみから、カメラの運動と対象物の形状復元を同時におこなうことができるが、本研究で利用するには、以下のような問題点がある。

- 多数の画像から得られた多数の特徴点の推移情報を、一括して処理するため、オクルージョン(点の隠れ)が発生する場合には処理が複雑になる。形状モデリングの逐次フェーズでは、物体の周囲を移動するカメラから取得した画像において、特徴点がかなり頻繁に入れ替わる。このため、因子分解法は初期フェーズでのみ利用している。
- 投影モデルを線形に近似することで分解をおこなうため、復元結果に誤差が含まれる。

非線形最小化

モーションステレオの際、カメラの位置・姿勢が正確でなかった場合の特徴点の復元結果や、因子分解法で求められたカメラ運動と物体の形状情報は、誤差を含んでいる。本研究においては、これらの誤差を補正するため、非線形最小化を利用する。非線形最小化は、初期値を入力し、勾配法による最適化をおこなうことで、評価関数を最小化する。

本研究で用いた評価関数を以下に示す。

$$C = \sum_{j=0}^2 \sum_{i=0}^P \left(\frac{\mathbf{r}_{xj}^T \mathbf{m}_{ij}}{\mathbf{r}_{zj}^T \mathbf{m}_{ij}} - \frac{X_i + t_{xj}}{Z_i + t_{zj}} \right)^2 + \sum_{j=0}^2 \sum_{i=0}^P \left(\frac{\mathbf{r}_{yj}^T \mathbf{m}_{ij}}{\mathbf{r}_{zj}^T \mathbf{m}_{ij}} - \frac{Y_i + t_{yj}}{Z_i + t_{zj}} \right)^2 \quad (6)$$

\mathbf{m}_{ij} は、j枚目の画像におけるi番目の特徴点の投影座標を表す。 \mathbf{r} は回転行列 \mathbf{R} の列ベクトルであり、 t_x, t_y, t_z はワールド座標系からカメラ位置への平行移動成分、 X, Y, Z は特徴点の3次元座標を表す。

本研究の実験システムは、カメラに俯角をつけ、それを搭載したロボットが床面上を走行するものである。可変なカメラパラメータは、カメラの俯角 θ 、カメラのパン角 ϕ 、ロボットの直進成分 d_x 、並進成分 d_y の合計4つとなる。

最小化には勾配法のひとつであるニュートン法を利用しており、計算時間は変数の数が多くなるほど増大する。初期フェーズでは、カメラパラメータ4個と特徴点の数 $\times 3$ 個を変数として、カメラの位置・姿勢と特徴点の3次元位置を同時に補正するために非線形最小化をおこなう。逐次フェーズでは、カメラパラメータ4個のみを変数とし、カメラの位置・姿勢の推定に利用する。