# ロボットハンドによって操作される物体の 視触覚情報に基づく位置姿勢計測

亀崎 康之 (九州大学) 小川原 光一 (九州大学) 倉爪 亮 (九州大学) 長谷川 勉 (九州大学)

# Estimation Method of the pose of an Object Manipulated by a Multi-fingered Robot-hand using Visual and Force Information

# \*Yasuyuki KAMEZAKI (Kyushu Univ.),Koichi OGAWARA (Kyushu Univ.), Ryo KURAZUME (Kyushu Univ.),Tsutomu HASEGAWA (Kyushu Univ.)

**Abstract**— This paper presents an estimation method of the pose of an object manipulated by a multifingered robot-hand using visual and force information simultaneously. In this method, we assume that the 3D shape of the object is known and estimate the pose of the object by aligning the 3D object model with the observed data. We use 3 different measures: (1) 3D point data estimated by a stereo-vision, (2) image edges, (3) contact position on the fingertip of the robot-hand. The pose is estimated by minimizing the weighted sum of the distances between the object model and each measure by a gradient descent method.

Key Words: Pose Estimation, Stereo Vision, Tactile Sensing, Sensor Fusion, Multi-Fingered Robot-Hand

# 1. はじめに

人間の手の構造を模した多関節多指ハンドは,器用 かつ柔軟な物体操作が実現可能な高機能ロボットハン ドである.しかし操作中の物体の位置姿勢がわからな ければ,多関節多指ハンドに期待される高度な物体操 作の実現は困難である.物体の位置姿勢計測手法とし て,力覚センサを用いてハンド指先と物体の接触位置 を観測し,接触点の変位量から物体の位置姿勢を推定 する方法がある.しかしこの手法では,物体と指先の 間に滑りが起きた場合に対応できない.そのため,内 界センサに加えて何らかの外界センサを用いた計測が 不可欠である.

内界センサと外界センサを併用する研究として,人 間の視触覚系との類似性から、ロボット搭載センサと カメラ画像を併用するアプローチが従来用いられてき た.長谷川ら [1] は,ステレオカメラを用いて観測し た物体表面の3次元位置と指先接触センサによって求 めた指先接触位置から物体の位置姿勢を計測する手法 を提案した.しかし,投光機を用いて特徴点となる模 様を物体表面上に投影する必要があり,明るい環境に おける計測や投影光と同色の物体の計測には使用でき ない.また横小路ら [2] は,単眼カメラを使用し,物体 表面に取り付けたマーカの画像上での変位量と接触位 置の変位量から物体の位置姿勢を計測する手法を提案 した.しかし,物体表面にマーカを設置する必要があ り,マーカの付いていない物体に対しては使用できな い.また,物体の高速な操りには高速な視触覚フィー ドバックが欠かせないが,画像処理の計算量は一般に 大きいため, [1][2] ではいずれも画像処理を高速に行う ために専用ハードウェアを使用していた.近年になっ て,汎用 PC に搭載されている GPU を用いて画像処 理を高速に実現する手法が注目を浴びており, ロボッ トビジョンの分野でも応用され始めている[3].

そこで本稿では,視覚処理に特別な準備や装置を必 要としない,視触覚情報に基づく物体の位置姿勢計測 手法を提案する.提案手法では物体の形状と初期姿勢 が既知であることを仮定し,3次元物体形状モデルと 観測データの位置合わせを毎時刻繰り返すことで物体 の位置姿勢を計測する.観測データとして,(1)ステレ オ視によって計測される3次元距離データ,(2)画像上 のエッジ,(3)指先接触位置の3種類の異なる情報を使 用する.これら3種類の観測データと物体モデル表面 との距離を定義し,その重み付き和を勾配法によって 最小化することにより物体の位置姿勢を求める.

#### 2. システム概要

本研究が対象とするシステムの概要を表した図を Fig. 1 に示す.ロボットハンドによって把持・操作される対 象物体を,4台のカメラによって構成される視覚と,指 先に1つずつ配置された6軸力覚センサを用いて観察 し,位置姿勢を計測する.ロボットハンドとカメラは 正確にキャリプレーションされているものとする.ま た,指先と物体間は点接触をしていると仮定する.



Fig.1 システム概要

## 2.1 ロボットハンド

本研究で使用するロボットハンドは,4関節の指1本, 3 関節の指2本の計3本で構成される.これら指の関 節は,それぞれ人間の拇指,示指,中指の自由度を参 考にして配置した.4 関節の指が拇指,3 関節の指が示 指と中指に相当する.各指の指先には6軸の力覚セン サが装着されており,指先にかかる力から指先座標系 における物体との接触位置を求めることができる.ま た,各指の関節には回転式のエンコーダが装着されて おり,各時刻におけるロボットハンドの姿勢を機械的 に求めることができる.さらに,ロボットハンドとカ メラがキャリブレーションされているため,カメラ座 標系における接触位置を計算することができる.

## 2.2 ステレオビジョンシステム

本研究で使用するビジョンシステムは4台のカメラ (PointGrey 社製 Flea2)によって構築される.ロボッ トが自身のハンドを見ながら物体を操作する状況を想 定し,400[mm]先の空間解像度が5[mm/pixel]以下に なるように設計した.水平方向と鉛直方向の解像度は 1.67[mm/pixel],光軸方向の解像度が4.53[mm/pixel] となっている.また,ステレオ演算の際にサブピクセ ル推定を導入することにより,実際はサブピクセル精 度で3次元計測を行うことができる.

#### 3. 物体の位置姿勢計測

提案手法では物体の形状と初期姿勢が既知であることを仮定し,物体モデルと観測データの位置合わせを 毎時刻繰り返すことによって物体の位置姿勢を計測する,使用する観測データは,

- ステレオビジョンによって計測した3次元点群
- 画像エッジ
- 指先接触点

の3つである.それぞれの観測データと物体モデルと の誤差を,物体モデルの位置姿勢パラメータqの関数  $E_{3d}(\mathbf{q})$ , $E_{2d}(\mathbf{q})$ , $E_c(\mathbf{q})$ として定義したとき,その重 み付き和 $E(\mathbf{q})$ を最小化するモデルの位置姿勢 $\mathbf{q}_{opt}$ を 求める.

$$\mathbf{q}_{opt} = \arg\min_{\mathbf{q}} E(\mathbf{q}) \tag{1}$$

$$E(\mathbf{q}) = w_{3d} E_{3d}(\mathbf{q}) + w_{2d} E_{2d}(\mathbf{q}) + w_c E_c(\mathbf{q}) \qquad (2)$$

以降の節では,それぞれの誤差関数の定義と,それを 用いた物体の位置姿勢計測手法について説明する.

## 3.1 誤差関数の定義

本節では,誤差関数 $E_{3d}(\mathbf{q})$ , $E_{2d}(\mathbf{q})$ , $E_c(\mathbf{q})$ の定義 についてそれぞれ説明する.

## 3.1.1 3次元点群に対する誤差関数



**Fig.2**3次元点群とモデルの誤差

ステレオビジョンによって計測した3次元点群と物 体モデル表面の点との対応を考える.物体の大まかな 位置姿勢は前時刻の推定結果より既知である.このと き, Fig.2 に示すように,前時刻における物体モデル 表面の点 $m_k$ とそれに対応する3次元点 $r_k$ との距離を 誤差関数と定義する.具体的には,モデル表面の点の うち,カメラから可視である点全てに対して推定誤差 を求め,その二乗和を誤差関数 $E_{3d}$ として以下のよう に定義する.

$$E_{3d}(\mathbf{q}) = \sum_{k \in N_{3d}} \| m_k(\mathbf{q}) - r_k \|^2$$
(3)

ただし, N<sub>3d</sub> はモデル表面の点のうち, カメラから可 視である点の集合である.

#### 3.1.2 画像エッジに対する誤差関数



Fig.3 画像エッジとモデルの誤差

画像平面に投影したモデルのエッジと画像エッジとの対応を考える.Fig.3に示すように,前時刻における物体モデルのエッジ $m_k$ を画像平面に投影し $mi_k$ とする.次に,画像平面上で $mi_k$ と対応する画像エッジ $ri_k$ を求め,カメラ中心と $ri_k$ を結ぶ直線上で $m_k$ との距離が最小となる点を $r_k$ とする.このとき $m_k$ と $r_k$ の距離を推定誤差として定義する.具体的には,モデルエッジのうちカメラから可視であるエッジ上の点すべてに対して推定誤差を求め,その二乗和を誤差関数 $E_{2d}$ として以下のように定義する.

$$E_{2d}(\mathbf{q}) = \sum_{k \in N_{2d}} \| m_k(\mathbf{q}) - r_k \|^2$$
(4)

ただし, N<sub>2d</sub> はモデルエッジ上の点のうち, カメラか ら可視である点の集合である.

#### 3.1.3 指先接触点に対する誤差関数



Fig.4 指先接触位置とモデルの誤差

ロボットハンドの指先と物体の接触点とモデル表面 の点との対応を考える.Fig.4に示すように,指先接触 位置 $r_k$ に対応する前時刻のモデル表面の点 $m_k$ を求め る.このとき $r_k$ と $m_k$ の距離を推定誤差と定義する. 全ての指先接触点に対して推定誤差を求め,その二乗和を誤差関数 *E<sub>c</sub>*として以下のように定義する.

$$E_{c}(\mathbf{q}) = \sum_{k \in N_{c}} \| m_{k}(\mathbf{q}) - r_{k} \|^{2}$$
 (5)

ただし, N<sub>c</sub> は指先接触点の集合である.

#### 3.2 ICP 法とロバスト推定に基づく位置姿勢計測

本節では,前節までに定義した誤差関数 $E(\mathbf{q})$ を用 いて対象物体の位置姿勢を計測する手法について説明 する.式(2)の右辺の項のうち $E_{3d}(\mathbf{q}) \geq E_{2d}(\mathbf{q})$ の項 は,画像処理による推定誤差により外れ値やデータの 抜けの頻度が高くなる.一般に誤差の分布が正規分布 で近似できない場合に最小二乗法を適用すると,式の 最小化が不安定になり妥当な位置姿勢が求まらない場 合がある.そこで,外れ値やデータ欠損の影響を軽減 するために,提案手法では[4]に習いロバスト推定であ る M 推定法を導入する.誤差分布はローレンツ分布で 近似し,関数 $\rho(z)$ を用いて式(3),(4)を以下のように 書き直す.

$$E_{3d}(\mathbf{q}) = \sum_{k \in N_{3d}} \rho(\| m_k(\mathbf{q}) - r_k \|^2)$$
(6)

$$E_{2d}(\mathbf{q}) = \sum_{k \in N_{2d}} \rho(\| m_k(\mathbf{q}) - r_k \|^2)$$
(7)

$$\rho(z) = \log\left(1 + \frac{1}{2}\left(\frac{z}{\sigma_z}\right)^2\right) \tag{8}$$

これまでに定義した誤差関数を用いて,勾配法の一 種である ICP(Iterative Closest Point)法により物体の 位置姿勢を計測する.具体的な処理の流れを Fig.5 に 示す.



Fig.5 ICP 法に基づく姿勢計測の流れ

位置姿勢計測システムは,ステレオビジョンによって計測した3次元点群データと,ビジョンシステムの中央のカメラの画像から抽出したエッジを入力とする.入力を受け取ると, $E_{3d}(\mathbf{q}), E_{2d}(\mathbf{q})$ の計算に必要なモデルと観測データの対応点探索を行う.このとき効率よく対応点探索を行う方法として,Kd-Treeなどの探索木を使用する方法が考えられる.しかし探索木を用いる方法は,探索の際の計算量は非常に小さいが,木を構築する際に大きな計算量が必要となる.そこで提

案手法では,画像平面上の近傍画素から探索する方法 を採用した.この方法では,探索の際の計算量は探索 木を用いる方法に比べ若干大きいが,木の構築にかか る計算量を削減することができる.また今後,対応点 探索の分散処理化及び GPU への実装によりさらなる 高速化も期待できる.一方, $E_c(\mathbf{q})$ の計算に必要な指 先接触点位置に対応するモデル表面点の探索には全探 索を採用した.これは,物体の裏側と接触している場 合などモデル表面上の指先接触位置が画像上で不可視 となる場合があることと,観測データ数が最大でも指 の本数と少ないため全探索でも計算量が大きくならな いことからである.

次に, $E(\mathbf{q})$ を $\mathbf{q}$ で偏微分し勾配ベクトルを求める. さらに,求めた勾配ベクトル方向に一次元探索を行い  $E(\mathbf{q})$ が最小となる点にモデルの位置姿勢を更新する. これを $E(\mathbf{q})$ が閾値以下になるか,繰り返し計算の回 数が閾値以上になるまで繰り返す.

#### 4. 実験

#### 4·1 実験の概要

本実験ではまず,3本指のロボットハンドによって操 作される対象物体をロボット搭載センサと視覚センサ を用いて物体を観察し,各時刻で得られる各センサの 出力を用いてオフラインで物体位置姿勢の計測を行っ た.計測に使用したデータ長は40フレームである.検 証した項目は,(i)物体を正しく追跡できているか,(ii) 計算時間,の2点である.ただし,(i)に関しては物体 位置姿勢の真値が未知であるため,画像上におけるモ デルと計測値とのずれと3次元空間におけるモデルと 計測値のずれを基に定性的に評価した.

また,実験では式 (2)の $w_{3d},w_{2d},w_c$ の値が計測結果 にどのような影響を与えるのか調べるために以下の4 通りのパラメータに対して実験を行った.

- 3次元点群データのみ  $(w_{3d} = 1, w_{2d} = 0, w_c = 0)$
- 画像エッジのみ (w<sub>3d</sub> = 0, w<sub>2d</sub> = 1, w<sub>c</sub> = 0)
- 3次元点群 + 画像エッジ (w<sub>3d</sub> = 1, w<sub>2d</sub> = 1, w<sub>c</sub> = 0)
- 3次元点群 + 画像エッジ + 指先接触点 (w<sub>3d</sub> = 1,w<sub>2d</sub> = 1,w<sub>c</sub> = 1)

#### 4·2 位置姿勢計測結果

実験の最終フレームにおける位置姿勢の計測結果を Fig.6 に示す.実験では前節の4通り全てのパラメータ において最終フレームまで物体を追跡することができた. ただし,3次元点群のみを使った計測では,画像平面に 水平な方向に対して計測誤差が大きくなり(Fig.6(a)), 画像エッジのみを用いた場合ではカメラの光軸方向に対 して計測誤差が大きくなった (Fig.6(d)). これら 2 つの 情報を組み合わせることで,画像平面の水平方向,光軸 方向共に精度よく計測することができた(**Fig.**6(e),(f)). さらに指先接触点の情報を加えた場合では,2つの視 覚情報を組み合わせた場合とそれほど大きな違いは見 られなかった (Fig.6(g),(h)). これは, 今回使用した実 験データが隠蔽のほとんどないデータで,視覚情報だ けで十分精度よく計測できたためであり、ハンドによ る隠蔽が頻発するようなデータでは,接触点データは 有効に使用できると考えられる.

また,物体の初期値が真値から大きくずれた場合や 途中で計測に失敗してしまった場合などに接触情報の 利用が有効であった.Fig.7 に初期値が大きくずれた

# RSJ2009AC1A3-07





(b) 3 次元点群のみ (3

計測詞

(d) 画像エッジのみ (3

次元)

次元)

(a) 3 次元点群のみ (画像)



(c) 画像エッジのみ (画像)





(e) 3 次元点群 + 画像エッジ (画像)





(f) 3 次元点群 + 画像

エッジ (3次元)

(g) 3 次元点群 + 画像エッジ + 指先接触点 (画像)

(h) 3 次元点群 + 画像エ ッジ + 指先接触点 (3 次 元)

Fig.6 計測結果 (Frame40)

### 場合に,指先接触点位置の重心に物体重心位置を移動 することで,エラー回避を行う例を示す.

#### 4·3 計算時間

計測に要した計算時間は平均 65[ms/frame] であった.実験には Intel(R)Core2Duo 3.16GHz(CPU) と NVIDIA GeForce GTX285(GPU) を搭載した PC を 用いて処理を行った.また,使用した画像の大きさは 320 × 240[pixel] である.各処理における計算時間の 内訳を Table 1 に示す.計算量の多いステレオ演算は





(a) 初期状態(b) 計測結果Fig.7 指先接触点位置を用いたエラー回避

GPU上で実装することにより高速化を行っている.全ての処理の中で勾配法による位置姿勢計測の処理に最も大きな計算時間がかかった.これは勾配ベクトルの計算の際に対応点探索が繰り返し行われることに起因する.提案手法では画像平面での探索を行っているため並列処理化が容易であり,今後 GPU に実装することでさらなる高速化が期待できる.

処理	計算時間
全体	65[ms/frame]
ステレオ演算 (GPU)	11[ms/frame]
エッジ抽出	2[ms/frame]
背景エッジ除去	9[ms/frame]
勾配法による位置姿勢計測	40[ms/frame]

Table 1 計算時間の内訳

#### 5. まとめ・今後の課題

視覚処理に特別な準備や装置を必要としない,視触覚 情報に基づく物体の位置姿勢計測手法を提案した.(1) ステレオカメラをによって計測される3次元距離デー タ,(2)画像上のエッジ,(3)指先接触位置の3種類の観 測データと物体モデル表面との距離を定義し,その重み 付き和を勾配法によって最小化することで物体の位置姿 勢を求めた.システム全体の処理時間は65[ms/frame] であった.

今後は,画像上での対応点探索を GPU 上で実装し 高速化を図るとともに,視覚の隠蔽の多いデータに対 する触覚情報の有効性を検証する.

- [1] 長谷川,本田,桐木,松岡,"多関節多指ハンドのための操作対象の位置と姿勢のセンシング-視覚情報と接触情報の融合-",電気学会論文誌C(電子・情報・システム部門誌),vol.118, no.9, pp.1340-1346, 1998.
- [2] 横小路,坂本,吉川,"カメラ画像を併用した多指ハンドにより操られる物体の位置姿勢推定法-ソフトフィンガー型ハンドによる操り制御への適用-",日本ロボット学会誌,vol.18,No3,pp.401-410,2000.
- [3] P.Michel, J.Chestnutt, S.Kagami, K.Nishiwaki, J.Kuffner, and T.Kanade, "GPU-accelerated Real-Time 3D Tracking for Humanoid Locomotion and Stair Climbing,"IEEE/RSJ Internationl Conference on Intelligent Robots and Systems, pp. 463-469, 2007.
- [4] Mark Damon Wheeler, "Automatic Modeling and Localization for Object Recognition", PhD thesis, CMU, 1996.