

## 「実演に基づく教示」におけるロボットビジョンの役割とその応用

小川原 光一, 木村 浩<sup>†</sup>, 池内 克史<sup>††</sup>

科学技術振興事業団, 東京大学生産技術研究所

<sup>†</sup> 電気通信大学

<sup>††</sup> 東京大学

〒153-8505 東京都目黒区駒場 4-6-1

東京大学生産技術研究所 第3部 池内研究室

03-5452-6242

ogawara@cvl.iis.u-tokyo.ac.jp

あらまし

実演に基づく教示とは, ロボットの動作をプログラミングする際に, 人間が実際に動作のお手本を実演し, それをロボットが観察することによって動作の獲得を図るものである. 本稿では, 実演教示におけるロボットビジョンの役割について議論し, 手作業を獲得する際に重要となる物体及び手形状の認識手法として, 筆者らが使用しているモデルベースの手法を説明する. また, この結果の応用として, 観測に伴う誤差を自動的に修正する方法, および複数の観察データから作業に本質的な箇所を自動推定する手法について紹介する. 最後に, まとめと今後の展望について述べる.

キーワード 実演教示 ロボットビジョン

## Role and application of robot vision in Learning from Observation paradigm

Koichi OGAWARA, Hiroshi KIMURA<sup>†</sup> and Katsushi IKEUCHI<sup>††</sup>

Japan Science and Technology Corporation, Institute of Industrial Science, The University of Tokyo

<sup>†</sup> University of Electro-Communications

<sup>††</sup> The University of Tokyo

### Abstract

In Learning from Observation paradigm, a human demonstrator performs a task in front of a robot system and the robot system automatically learns the intended task through observation. In this paper, the role of robot vision in observation process is discussed. Then, to understand the demonstrated task, some model-based recognition methods to estimate the pose of the objects and the shape of the hand are described. To extract high level representation of the subject task, 2 techniques are presented: one is to correct observation error in contact relations and the other is to estimate essential interactions to complete the task from multiple demonstrations. Finally, we conclude and refer to future directions.

**Keywords** Learning from Observation, Robot Vision

# 1 はじめに

近年、家庭やオフィス等の人間共存社会で行動することを目的とした共生型ロボットの開発が積極的に行われているが、このような多様かつ動的に変化する環境で使用されるロボットには、環境もしくはユーザのニーズに応じて新規の動作を適応的に獲得する能力が求められる。「実演に基づく教示」(以下実演教示)とは、ロボットに関する専門知識を持たないユーザが、人が人に対して動作を教えるのと同様のやり方でロボットに対して動作の教示を行うための手法であり、上述のロボットに対する簡便な動作教示方法として用いられることが期待される。本研究においては、対象とする実演教示動作を日常で一般に行われる手作業とする。このとき、ロボットはお手本動作をセンサによって観察し、観察データから動作を復元可能な何らかの抽象表現を生成する。この抽象表現は、教示者の手の形状・位置姿勢、操作物体の位置・姿勢・力といったコンピュータビジョンのアルゴリズムを直接適用して得られる低レベルの情報から、それを加工して得られる操作物体と環境との接触状態や特定動作の作業全体における重要性などの高レベルの情報も含め、タスクに応じて決定される。また、ロボットがこの抽象表現を元に動作を再現する場合にも、抽象表現から逸脱しないために自身の行為を観察し、動作もしくは抽象表現自体を補正する必要がある。

観察を行うために用いるセンサとして、接触型のセンサと非接触型のセンサがある。接触型のセンサとしては、データグローブや位置・姿勢・力センサ等が主に用いられ、上記の低レベルの情報を精度よく獲得することが可能である反面 [1, 2], 煩雑なセットアップや、これによって教示動作そのものが拘束されることが教示者の負担となる。

一方非接触型センサ、つまりロボットビジョンとしては、CCD カメラやレンジセンサなどが用いられるが、レンジセンサについては直接距離の情報が得られるものの [3], 一回の計測に時間がかかることから連続的に観察する必要のある動作には適していないため、本稿ではロボットビジョンとして主に CCD カメラを用いた場合について議論を進める。

以降の章では、まず第 2 章においてその役割を議論する。第 3 章では、対象の 3 次元幾何モデルが既知である場合に、教示動作中の環境物体・操作物体・手の位置・姿勢・形状を推定するモデルベースの手法について述べる。第 4 章では、第 3 章の結果を利用して作業に関する高レベルの表現を得る方法として、組立て作業における観察誤差を補償する方法と、複数の教示動作を利用して作業に本質的な動作を事前知識無しに推

定する手法について説明する。最後に、第 5 章においてまとめと今後の課題について述べる。

## 2 実演教示におけるビジョンの役割

実演教示におけるビジョンの役割は、ロボットが教示動作を再現するために必要十分な情報を獲得することにある。この必要十分な情報とは、教示対象となる作業の種類や、必要とされる再現レベルによって異なる。

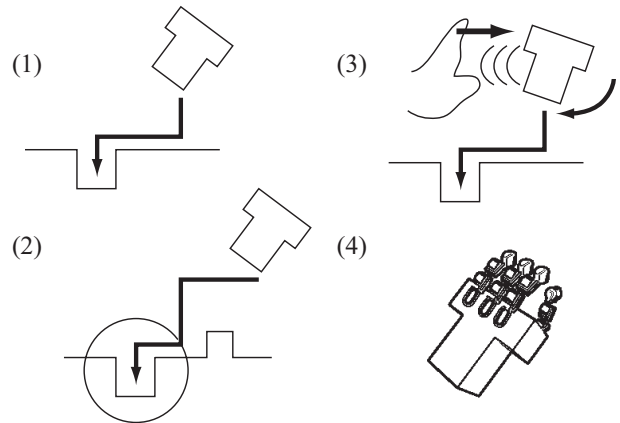


図 1: タスクの遂行に必要な情報

図 1(1) のようにペグをホールに入れる作業の場合、環境の変化がない場合には、ペグとホールとの接触状態の変化を観測し、同じ順序で接触状態の変化を生じさせるロボット動作を順に実行することによって作業の再現が可能になる。この場合、起りうる接触状態変化の数が有限個の場合には、抽象表現として必要な情報は、連続する接触状態変化を表す記号列となる。一方、図 1(2) のように実行時の環境が変化する場合には、ロボットはまず対象となるホールを環境のなかから探し出す必要がある。ホールの認識のための情報が必要がある。さらに図 1(3) のように、何らかの外乱によって操作物体が影響を受ける可能性がある場合には、ロボットは操作物体の形状・位置姿勢に関する情報を取得しておき、実行時に自身の動作を観察し実時間で動作を補正する必要がある [6]。また、図 1(4) のように多指ハンドを備えたロボットに対して物体のマニピュレーションを教示する場合には、把持のマッピングのために教示者の手の形状を認識する必要がある。

このように、目的とするタスク・環境条件によって必要な情報が異なるため、Task Oriented Vision[4] と呼ばれるように、タスクに関する知識を利用して高速かつロバストな視覚システムを設計をすることが一般的であった。

國吉らは、対象の動作領域を平面に仮定することに

より、単眼の CCD カメラから作業者の指先や物体の境界エッジ等の特徴点をトラッキングし、それらの相対関係の変化を調べることによって作業の認識を行った [5]。池内らは、任意の凸物体同士の間と面との接触状態変化が有限の状態数で記述できることを利用し、視覚のみから面接触状態遷移列を獲得し同じ状態遷移をロボットハンドによって再現することにより組立作業の獲得と再現を行う Assembly Plan from Observation (APO) を提案した [3]。

また、ロボットと教示者が対面している場合、教示時のロボットからの見えと再現時のロボットからの見えが異なるため、再現に必要な情報がオクルージョン等の影響によって十分に得られない可能性がある。厳密には非接触型ではないが、横小路らは教示者の頭部に小型のステレオカメラを装着し、これを教示用の視覚として用いることでこの問題に対処する方法を提案している [7]。

一般に、撮像時の量子化誤差、物体認識時の推定誤差などにより、視覚から得られた情報は誤差を含んでいる。上記の研究では、視覚によって得られた対象物体の位置・姿勢等の情報をそのままロボットの動作指令に変換するか、位置の誤差を補償するために力の情報を利用していった。しかし、このままでは認識時の解析に不都合であり、この誤差を後段の処理に影響しない範囲に収める必要がある。この一環として、組立て作業の接触状態解析において、解析の前に接触に関する誤差を修正する方法を第 4 章で述べる。

また、環境中の複数の物体と操作物体とが連続的に相互作用を行う作業においては、上記のように一連の教示動作全体が動作を再現する上で必要なのではなく、離散的に現れる相互作用を異なる環境において再現可能なように個別にモデル化することが重要となる。そのための方法として、第 4 章において複数の観察データを相互比較し、教示作業全体からタスクに本質的な相互作用を自動抽出する方法について述べる。

### 3 視覚処理に基づく作業の解析

教示動作を理解する場合には、場面で何が行われているのかを認識する必要がある。本研究では手作業を対象としているため、操作物体・被操作物体の動きの情報、教示者の手の動き・形状に関する情報の 2 種類を視覚によって獲得することが重要になる。

本章では、操作物体の任意の 3 次元軌跡を追跡する方法及び、把持物体によって隠蔽された手形状を推定する手法について述べる。いずれも、対象の 3 次元形状モデルを既知とし、複数のカメラより対象の 3 次元

形状を復元し、3 次元形状モデルとのフィッティングを行うことによって推定を行う。

#### 3.1 同定手法

対象の 3 次元幾何モデル (3 角パッチモデル) が既知であり、また位置や姿勢が不明な対象の 3 次元形状が何らかの方法によって復元できる場合に、対象の位置・姿勢  $\langle q, t \rangle$  を適切な初期位置  $\langle q_0, t_0 \rangle$  を元にロバーストに推定する 3D Template Matching (3DTM) 法 [8] について説明する。ここで、 $t$  は位置を表す 3 次元ベクトルであり、 $q$  は姿勢を表す 4 次元の quaternion ベクトルである。

幾何モデルの  $i$  番目の頂点のモデル座標系での 3 次元位置を  $mp_i$  とし、距離データ中の対応する 3 次元点を  $rp_i$  とすると、両者の関係は変換式  $p = \langle q, t \rangle$  によって誤差を含んだ式 (1) の形で表される。ただし、演算  $q \cdot mp_i$  は、 $mp_i$  を 4 次元ベクトルである quaternion  $q$  に従って回転させることを意味する。

$$rp_i = q \cdot mp_i + t + \beta \quad (1)$$

ここで、 $\beta$  はランダムな 3 次元誤差であり、 $\beta$  がガウス分布に従うならば、 $p$  は最小二乗法により式 (2) を最小化することによって解くことができる。

$$E(p) = \sum \| q \cdot mp_i + t - rp_i \|^2 \quad (2)$$

しかし、実際の誤差分布は通常ガウス分布に従わないため、最小二乗法の場合外れ値の影響により同定処理が不安定になる。そのため、Wheeler は M 推定法を誤差分布の評価に用いた物体の同定手法を提案した [8]。M 推定法は最小二乗法の一般形であり、 $E = \sum_i \rho(z_i)$  と定式化される。この関数を、共役勾配法を用いた反復計算によって漸近的に局所最小値に落とすことにより  $p$  を推定する。ただし、 $\rho(z_i)$  は誤差  $z_i$  の任意の関数である。式 (3) で表される Lorentzian 関数の 2 乗を  $\rho(z)$  として用いることで、外れ値の影響を低減し同定性能を向上させることができる。これにより、対象物体の一部が隠蔽され不可視の場合にも、メッシュモデル上で正しく対応する距離データがない箇所は外れ値として共役勾配法への寄与が低減されるため、正確な同定が可能となる。

$$\psi(z) = \frac{\frac{z}{\sigma}}{1 + \frac{1}{2} \left(\frac{z}{\sigma}\right)^2} \cdots \text{Lorentzian weight function.} \quad (3)$$

ただし、 $\sigma$  は重みである。

### 3.2 物体認識

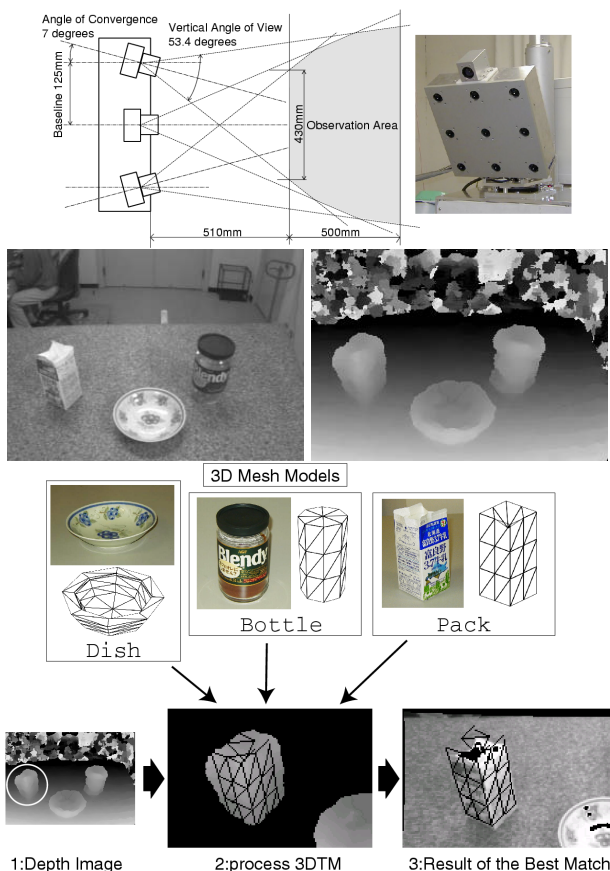


図 2: 多眼ステレオを利用した物体の同定

環境中の物体を認識する場合には、図 2 の多眼マルチベースラインステレオ装置を利用し、生成されるカラー画像及びディスパリティ画像より物体の抽出を行う。本研究ではテーブル上での手作業を仮定しているため、テーブル上が空の状態である背景画像をあらかじめ撮影しておき、距離データに対する背景差分法によってまず候補となる物体領域を抽出する。

次に、この各領域に対して、その重心を初期座標とし、上記の 3DTM によって各 3 次元モデルの位置合せを行う。初期姿勢については、異なる 4 方向の姿勢を与え、同定結果の最もよかったものを選択する。このとき、ステレオによって得られる距離データは  $2\frac{1}{2}$  次元の情報であるため、同定の際も幾何モデルの 3 角パッチのうち法線が視点側を向いているものだけを使う必要がある。位置合せが収束した後に、M 推定を用いない純粋な二乗誤差の平均を出し、誤差のもっとも小さなモデルを対象物体とする。また、物体の色相ヒストグラムをモデルとして持っておき、カラー画像上の候補領域の色相ヒストグラムとの差が閾値以上の場合には、同定を行う前に候補モデルを棄却することで性能

の向上を図っている [6]。

### 3.3 軌跡推定



図 3: 操作物体の 3 次元追跡

ステレオ装置により、最大 30fps の周期で連続するカラー画像と距離データを得ることが可能である。対象物体の時刻 0 での位置は上記の認識手法によって既知であるため、教示者が操作する物体の軌跡  $\langle q(t), t(t) \rangle$  を獲得するためには、 $\langle q(t-1), t(t-1) \rangle$  を初期値として、各物体毎に上述の 3DTM を連続的に適用することによって図 3 のように求められる。

### 3.4 把持形態の認識 [9]

手作業においては、必然的に物体を把持する場面が多くなるが、従来提案されてきた手の形状認識手法はジェスチャ認識などを念頭に開発されているため、把持を考慮していないものが大半であった。把持を伴う場合の問題点として、手の姿勢によってはカメラ画像上で手が把持物体により隠蔽されるため、この場合欠損データから推定することになることから推定精度が低くなる点があげられる。

そこで、我々は指を積極的に動かす「操り」を伴わない把持を対象とし、把持期間中手の形状は変化しないが手の姿勢は変化するという仮定の下で、連続する時系列データから  $N$  枚のサンプルを取りだし、この中で欠損データからの影響をうまく低減することによって、結果的に欠損の少ないデータのみを利用して手の形状を推定する手法を提案する。

具体的には、光軸が直交するように配置した 3 台の赤外線カメラを用い、把持を行いつつ姿勢が時間軸上連続に変化する手の連続画像を獲得する。次に、時系列データの中から一様な時間間隔で  $N$  セットの画像を抽出し、各セットについて図 4 に示すように 3 枚の画像から視体積交差法によって手のボリューム表現を復元する。一方手の幾何モデルについては、手の表面形状を表すメッシュモデル(皮膚)を 2.5 関節のリンク構造(骨)にアタッチし、関節角を指定することによって自由変形する手の形状モデル (Hand articulated model) を作成する。次に、式 (4) に示すように、3DTM を開



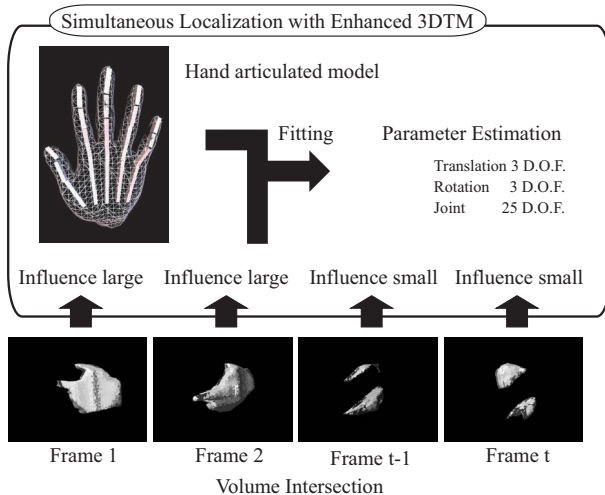


図 4: 把持形態の推定

節物体のフィッティングに適用できるように拡張した手法 (Enhanced 3DTM) を用いて、手の形状モデルの復元データに対するフィッティングを行い、姿勢と関節角の推定を行う。

$$E(\mathbf{p}, \boldsymbol{\theta}) = \sum_i \rho \left( \left\| R\mathbf{m}_i(\boldsymbol{\theta}) + \mathbf{t} - \mathbf{r}_j \right\|^2 \right) \quad (4)$$

$$\begin{pmatrix} \mathbf{m}_i(\boldsymbol{\theta}) \\ 1 \end{pmatrix} = T_l(\boldsymbol{\theta}) \begin{pmatrix} \mathbf{m}_i \\ 1 \end{pmatrix}$$

ここで、 $T_l$  は  $l$  番目の関節角を変数にとり、モデルのローカル座標系でメッシュの頂点  $m_i$  の位置を変換する  $4 \times 4$  の同次行列を表す。

このフィッティングの際に、欠損画像によって復元されたボリュームデータは大半が誤対応を引き起こす。そこで、 $N$  個の復元データ全てを同時に用いて関節角の収束計算を行うことで、M推定により誤対応の寄与を正しい対応からの寄与よりも相対的に低くすることができるため、結果的に欠損データを棄却することによって手の形状推定が可能になる。

図 4 の場合は、円筒形の物体を把持した状態で手を回転させた時系列データであるが、時刻  $t-1$  と  $t$  の復元データはオクルージョンによって指の大部分が欠損している。この場合、モデルの指に正しく対応する復元データがなく誤対応となるため、M推定によって影響が軽減される (Influence small)。シミュレータによる定量的な評価では、関節角度の推定誤差は約 15 度であった。これは、復元データの条件がよい場合でも不完全な Visual Hull にしかならないことを考えると十分な結果であり、またこの程度の誤差であれば手の形状判断の指標になりうる。また実環境における実験においても、よい推定値が得られることを確認している。

## 4 視覚処理データの応用

前章では、視覚から得られるデータに対して直接コンピュータビジョンの手法を適用した結果について述べたが、本章ではこれらの情報を加工し、作業を解析する上で有用な高位の情報を得る試みを 2 つ簡単に紹介する。

### 4.1 接触状態の誤差修正 [10]

第 3 章で述べた方法を用いて操作物体の 3 次元軌跡や環境物体の位置を計算した場合、これらは一般に誤差を含んでいるため、接触を伴う作業においては物体同士の埋め込みや離脱が生じてしまう。そのため、図 1 の組立て作業のように拘束が増える方向で作業が進むタスクの接触状態を解析する場合、このままのデータでは正確な接触状態を推定することはできない。

このとき、対象とする物体の幾何モデルが与えられている場合には、操作物体の位置・姿勢の誤差のみを考慮すればよくなるため、次の 2 段階の方法によって誤差を解消することができる。

1. 接触が生じている個所の候補を検出する
2. 全ての接触候補を利用して、妥当な接触状態を推定する

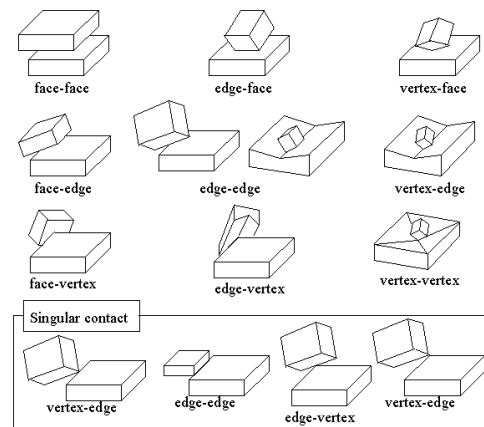


図 5: Nine types of contact elements

第 1 段階では、接触要素として Vertex-Face, Face-Vertex, Edge-Edge の 3 種類を考え、画像より推定された位置の情報から、操作物体と環境物体との間の全ての接触要素の組み合わせについて、これらの距離がある閾値より小さい場合を接触候補とする。多面体の場合には、図 5 に示すように 9 種類の接触関係がありうるが、これらはこの 3 つの接触要素の組み合わせで記述できるため、この 3 つが必要十分である。

第2段階では、各接触要素毎に誤差量の解析式が出るが、これらをテイラー展開によって線形形式に変換すると、いずれもの接触要素についても式(5)の形で表されることが分かる。

$$\Delta e = c - [a^T b^T] \begin{bmatrix} T \\ r \end{bmatrix} \quad (5)$$

ここで、 $a, b, c$  は各接触要素毎に決まる定数であり、 $T$  と  $r$  はそれぞれ操作物体の並進ベクトルと回転を表すオイラー角のベクトルである。

次に、最小二乗法を用いて式(6)を解くことによって、誤差を最小にする操作物体の位置姿勢を求めることができる。

$$\frac{\partial}{\partial q} \sum (\Delta e)^2 = 0 \quad (6)$$

## 4.2 必須動作の推定 [11]

従来の多くの解析手法と同様に一回の観察結果のみを利用して解析を行う場合、実演された教示動作の観察結果から作業の遂行に本質的な相互作用を推定することは困難である。これは、相互作用のない全身運動の獲得の場合 [12, 13] や、プリミティブ自体はあらかじめ設計され相互作用も明らかである場合 [3, 14, 5] には問題にならないが、一般の手作業のプリミティブを獲得する場合には (1) どの相互作用が作業の遂行に必須であるのか、(2) 把持物体とどの環境物体が相互作用を行っているのかの2点についてあいまい性があり、一回の観察結果からでは事前の知識なしにプリミティブを確定することはできない。

そこで、本質的には同一の作業を示す  $N$  個の異なる教示動作の観察結果を統合し、すべての教示動作に共通する相互作用を抽出することで作業に必須の動作と相互作用を行う物体ペアを同時に推定する手法を提案する。

まず、本質的には同一の作業を表す教示動作を複数  $N$  回実行しこれを記録する。ここで本質的に同一とは、作業の目的を達成するために必須の把持物体・環境物体間の相互作用が決まった形態・順序で現れるものと定義する。このとき、必須ではない相互作用が任意の順番で現れることは許す。すると、各教示動作は必須相互作用を内包した相互作用の連続形として表現することができる。

前述の理由から単一の観察結果には解析する上であいまい性が存在するため、そこから必須相互作用のみを抽出することは困難である。そこで、教示動作を複数回行う際に、必須相互作用に影響を与えずに環境も

しくは教示動作自体を毎回若干変化させる。すると、必須相互作用は同じであるが、それ以外の相互作用については教示動作によって異なる観察結果を得ることができる。

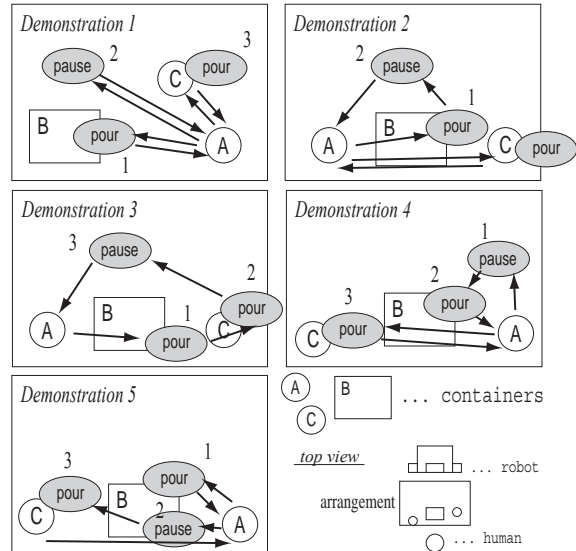


図 6: 複数の教示動作

図6は、対象作業として「注ぎ動作」を選択し、5回教示を行った例である。この例では、本質的な動作は (1) 容器 A の中身を容器 B に注ぎ、(2) 次に容器 A の中身を容器 C に注ぐことである。個々の教示動作の違いは、物体の配置や必須動作と関係のない手の動きにある。

次に、複数の観察結果を相互比較し、すべての観察結果に共通する相互作用を抽出することで必須の相互作用を推定する。相互比較を行う方法として、多次元 DP (Dynamic Programming) Matching [16] を用いる。

各観測結果を要素配列  $Seq_i (1 \leq i \leq N)$  の形で表現したときに、DP Matching を適用するには任意の要素の組合せ  $(Seq_1[index_1], \dots, Seq_N[index_N])$  の一致度を評価する関数  $E(index_1, \dots, index_N)$  を作成する必要がある。ここで、要素は検出された相互作用候補を指す。

$G(index_1, \dots, index_N)$  を、その要素の組合せに至るまでの評価値の重み付き和とすると、以下の漸化式によって最適パスつまり最適対応を計算することができる。すなわち、計算された最適パスが経由する要素の組合せが推定された必須相互作用である。

$$[s] \quad G(index_1, \dots, index_N) = E(index_1, \dots, index_N) \cdot \text{path}(ii, \text{index})/2 + \min\{\cup[G(ii_1, \dots, ii_N) + E(ii_1, \dots, ii_N) \cdot \text{path}(ii, \text{index})/2]\}$$

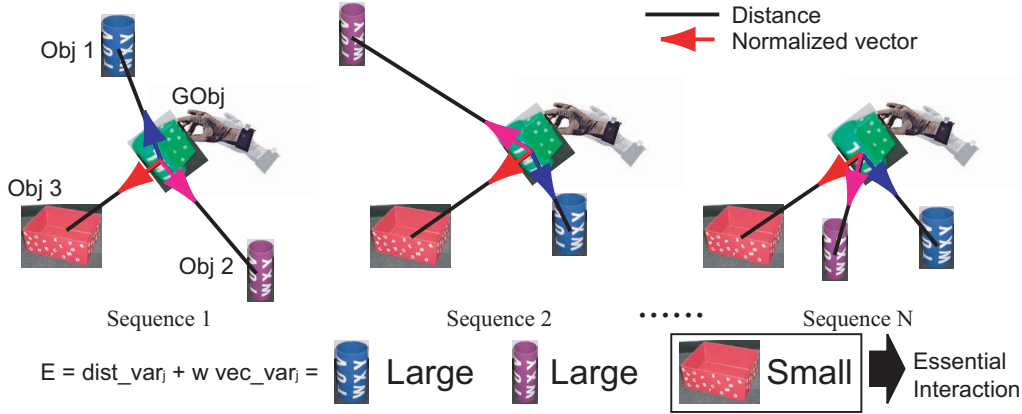


図 7: 必須相互作用の評価

$$\begin{aligned} \text{index}_1 - S \leq ii_1 < \text{index}_1, \dots, \\ \text{index}_N - S \leq ii_N < \text{index}_N. \end{aligned}$$

$$E = \min_j \{ \text{dist\_dev}_j + w \cdot \text{vec\_dev}_j \}. \quad (7)$$

ただし,  $\text{path}(ii, \text{index})$  は現在の格子点  $\text{index}$  から探索対象の格子点  $ii$  までの距離 ( $\sqrt{\sum_{k=1}^N (\text{index}_k - ii_k)^2}$ ) を表し, 探索範囲内の連続した必須相互作用が縮退して一つの必須相互作用として扱われることを防ぐ役割を持つ. また, (1) ある要素が 2 回以上他の配列の要素と一致することを許さない, (2) 探索範囲を  $N^S$  の格子とする, との 2 つの拘束条件を設定する. 前者は, 必須相互作用が他の必須相互作用と重なって検出されないための条件である. 後者の  $S$  は, 必須相互作用間に入りうる必須ではない相互作用の数を規定するものである.  $S - 1$  が連続する必須ではない相互作用の数よりも少ないと, 必須ではない相互作用が必須であると誤検出され,  $S$  を大きくした場合には計算コストが増大する. 各教示動作を見たときに最大の相互作用候補数に合わせるのが最も安全である.

ただし  $w$  は重みを示す定数であり, 本研究では両者のバランスがとれるように経験的に定めた.

図 7 の例では, 把持物体と  $Obj_3$  との関係 ( $GObj, Obj_3$ ) はすべての教示動作を通してほぼ同一である. そのため, 評価関数  $E$  ではこの関係 ( $GObj, Obj_3$ ) を表す重み付き分散和  $dev_3$  が他の関係のもの比べて最小となるため, この値を評価値として返す. この関係 ( $GObj, Obj_3$ ) が相互作用候補となり, この評価値が小さい程これが必須動作である確率が高くなるが, 最終的には DP マッチングの結果によってどの相互作用候補が必須であるのかが決定される. 逆に, もし評価関数  $E$  が小さい値を返さない場合には, ( $GObj, Obj_i$ ) ( $1 \leq i \leq 3$ ) はいずれも必須相互作用ではないと予想することができる.

#### 4.2.1 DP Matching における評価関数

本稿では, 把持物体と環境物体との相互作用の検出を目的としており, ある要素の組合せにおいて把持物体とある環境物体との相対関係がすべての要素に渡って一致しているときに, その環境物体と把持物体が必須の相互作用をしている (評価関数  $E(\text{index}_1, \dots, \text{index}_N)$  が低い値を返す) と見做す.

ここで相対関係とは, 図 7 に示すように把持物体  $GObj$  からある環境物体  $Obj_j$  までの距離  $\text{dist}_i(GObj, Obj_j)$  及び, 把持物体からその環境物体へ向けた正規化ベクトル  $\text{vec}_i(GObj, Obj_j)$  によって規定され, 両者の標準偏差の重み付和の最小値をもって評価値とする.

評価関数  $E$  は式 (7) のように表される.

#### 4.2.2 ロボットによる動作の再現



図 8: 作業の再現

図 8 に再現されたロボット動作の連続画像を示す. 実現の際には, 物体の配置は教示のときの配置とは異なっているが, 把持物体の軌跡が対象物座標系で教示のときと同一になるようにロボットアームを制御することによって同じ作業を再現することが可能になる. 各必須相互作用間を結ぶ動作は, 前の相互作用の終了位置と次の相互作用の開始位置を直線で結ぶ軌道によって

生成された。

## 5 まとめ

本稿では、実演教示におけるビジョンの役割について議論し、コンピュータビジョンのアルゴリズムによって対象の位置や形状を推定する低レベルの情報獲得法、およびこれらを加工することによる高レベルの情報獲得手法について筆者らの研究例を紹介した。

多くの実演教示に関する研究では、対象とするタスクを決定した上で、その中で起りうる全ての事象  $E_i$  を有限数定義しそれぞれに記号  $S_i$  をつける。観察においては、教示動作がどの事象と適合するのかを解析し、対応する記号の連続系として記述を行う。この場合、ロボットの扱っている作業は、最初の起りうる事象を設計する段階で決まってしまう、それを環境に応じて拡張することができない。第4章で述べた本質的な動作の自動推定の例は、これを打破する研究の一環でもあり、教示の類似性のみから作業に固有の動作群を抽出し、これらが独立な動作となるようにクラスタリング [17] することによって、環境に適応した作業のプリミティブ、つまりビルディングブロックを生成することが可能になる。

今後は、事象に関するプリミティブの動的な獲得に加えて、現在は人間が与えている物体の形状モデルについても自動的に獲得する方向に進みたい。この時、視覚のみによる抽出ではなく、ロボットがセンサを分布したハンドによって未知の物体をハンドリングすることにより、視覚と体性感覚とを同時に取り込みこれらを結合しモデル化することが本質的に重要であると考えている。

教示のために視覚以外のモダリティ、例えば音声等を同時に利用することも重要である。例えば、複数教示動作の統合例では、教示者が必須動作を行うタイミングで発声による合図を送れば、動作の推定がずいぶん楽になり、さらに上記の方法の制限であった抽出可能な動作条件が緩和される。現在、これに取り組んでいるところであり、よい結果を得ている。

また、観測された作業の情報から、作業内での動作や物体の動きに関するタスクに固有の文脈が生成され、逆にそのタスクにおいてはこの文脈が視覚処理の拘束条件になると考えられる。このように、視覚処理が独立して働くのではなく、他のセンサの情報もしくはより上位の情報が視覚処理の内容と相互作用を行う仕組みが、効率的な作業理解のために必要であろう。

## 参考文献

- [1] K. Ogawara, S. Iba, T. Tanuki, H. Kimura, and K. Ikeuchi. Recognition of human task by attention point analysis. In *Int. Conference on Intelligent Robot and Systems*, Vol. 3, pp. 2121–2126, 2000.
- [2] 津田雅之, 高橋友一, 小方博之. 実演教示による組み立て作業モデルの生成. 日本ロボット学会誌, Vol. 18, No. 4, pp. 73–82, 2000.
- [3] K. Ikeuchi and T. Suehiro. Toward an assembly plan from observation part i: Task recognition with polyhedral objects. *IEEE Trans. Robotics and Automation*, Vol. 10, No. 3, pp. 368–384, 1994.
- [4] K. Ikeuchi and M. Herbert. Task oriented vision. In *Int. conf. on Intelligent Robots and Systems*, pp. 2187–2194, 1992.
- [5] Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching. *IEEE Trans. Robotics and Automation*, Vol. 10, No. 6, pp. 799–822, 1994.
- [6] K. Ogawara, H. Kimura, and K. Ikeuchi. Refining hand-action models through repeated observations of human and robot behavior by combined template matching. In *Int. Conference on Intelligent Robot and Systems*, 2001.
- [7] 横小路泰義, 北岡佑輝, 吉川恒男. 動作者の視点からのカメラ画像によるモーションキャプチャとロボット作業教示への応用. 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol. 43, pp. 105–116, 2002.
- [8] M. D. Wheeler and K. Ikeuchi. Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 3, pp. 252–265, 1995.
- [9] 橋本謙多郎, 小川原光一, 高松淳, 木村浩, 池内克史. 赤外線画像を用いた視体積交差法による把持形体の認識. 情報処理学会研究報告 (CVIM 135), 2002.
- [10] J. Takamatsu, K. Ogawara, H. Kimura, and K. Ikeuchi. Correcting observation errors for assembly task recognition. In *Int. conf. on Intelligent Robots and Systems*, pp. 232–237, 2002.
- [11] 小川原光一, 高松淳, 木村浩, 池内克史. 複数教示動作の時系列上での統合に基づく人間作業のモデル化手法. 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol. 43, pp. 117–126, 2002.
- [12] T. Inamura, Y. Nakamura, H. Ezaki, and I. Toshima. Imitation and primitive symbol acquisition of humanoids by the integrated mimesis loop. In *Int. conf. on Robotics and Automation*, pp. 4208–4213, 2001.
- [13] A. Fod, M. Mataric, and O. C. Jenkins. Automated derivation of primitives for movement classification. In *IEEE-RAS Int. Conf. on Humanoid Robots 2000*, 2000.
- [14] 津田雅之, 小方博之, 南條義人. 実演教示による多角形ワーク組立のための局所的な接触状態遷移モデル群の生成. 日本ロボット学会誌, Vol. 18, No. 4, pp. 83–92, 2000.
- [15] B. Dufay and J. C. Latombe. An approach to automatic robot programming based on inductive learning. *Int. Journal of Robotics Research*, Vol. 3, No. 4, pp. 3–20, 1984.
- [16] G. Fuellen. A gentle guide to multiple alignment. *Complexity International*, Vol. 4, pp. 1–56, 1997. <http://www.csu.edu.au/ci/vol04/mulali/mulali.html>.
- [17] K. Ogawara, J. Takamatsu, H. Kimura, and K. Ikeuchi. Modeling manipulation interactions by hidden markov models. In *Int. Conference on Intelligent Robot and Systems*, 2002.