

# 9眼ステレオとデータグローブを用いた ロボットによる人間行動の認識と再現

小川原光一 射場総司 田貫富和 木村浩 池内克史

東京大学生産技術研究所 第3部

〒106-8558 東京都港区六本木7-22-1 Tel: 03-3401-1433

E-mail: {ogawara,ki}@iis.u-tokyo.ac.jp, iba+@cmu.edu,

tomikazu\_tanuki@komatsu.co.jp, hiroschi@kimura.is.uec.ac.jp

あらかし注視点 (Attention Point: AP) 解析 (以下 AP 解析) による人間行動モデルの構築法について述べる。AP 解析は2段階から成り、最初の段階では人間の行動を大まかに観察し、一連の行動の中からより詳しい解析を必要とする注視点 (AP) を発見する。第2段階では、同じ人間行動を再度観察し、今度は先程発見した注視点 (AP) 付近についてのみ時間をかけた重点的な解析を行う。こうすることで効率的に人間行動をモデル化することが可能となる。また本人間行動モデルは高度に抽象化されており、その抽象度を環境に応じて変化させることができるため、異なる環境への適用性が高い。

本稿では、データグローブと9眼ステレオビジョンを用いた AP 解析手法及び人間行動モデルの構築法について述べる。また実ロボットを用いた人間行動の認識及び異なる環境における再現実験を行うことで、本人間行動モデルの適用性について検証を行った結果について述べる。

## Recognition and Performance of Human Task with 9-eye Stereo Vision and Data Gloves

Koichi OGAWARA Soshi IBA Tomikazu TANUKI  
Hiroshi KIMURA Katsushi IKEUCHI

Institute of Industrial Science

University of Tokyo

7-22-1 Roppongi, Minato-ku, Tokyo JAPAN 106-8558 Tel: 03-3401-1433

E-mail: {ogawara,ki}@iis.u-tokyo.ac.jp, iba+@cmu.edu,

tomikazu\_tanuki@komatsu.co.jp, hiroschi@kimura.is.uec.ac.jp

**Abstract** This paper presents a novel method of constructing a human task model by attention point (AP) analysis. The AP analysis consists of two steps. At the first step, it broadly observes human task, constructs rough human task model and finds APs which require detailed analysis. Then at the second step, by applying time-consuming analysis on APs in the same human task, it can enhance the human task model. This human task model is highly abstracted and is able to change the degree of abstraction adapting to the environment so as to be applicable in a different environment.

We describe this method and its implementation using data gloves and a stereo vision system. We also show an experimental result in which a real robot observed a human task and performed the same human task successfully in a different environment using this model.

# 1 はじめに

ロボットが観察によって人間の行動を学習し、自動的に行動のレパートリーを増やすことができるようになれば、人間・ロボット共存社会においてロボットの適用範囲を大幅に拡大することが可能になる。

人間行動を獲得するためには、ロボットは何らかの人間行動モデルを構築する必要がある。この人間行動モデルは、ロボットが人間と同じ行動を再現したり、人間との協調作業を行う際に利用される。

これまで、視覚を用いたロボットによる人間行動学習の研究 [1, 2] や視覚に基づく人間・ロボット協調行動の研究 [3] が行われてきたが、これらの研究においては人間の行動観察は1回のみであった。つまり、1回の観察で人間行動モデルを構築した後、再度人間行動の観察を繰り返し解析を重ねて人間行動モデルを拡充させようという試みは行われてこなかった。しかし1回の解析において、人間の行動全てに渡って詳細な解析を行うことは現実的ではなく、かといって全般におおまかな解析を行うのでは、詳細な解析を必要とする個所の解析が不十分となる。

本研究では人間行動の解析を2段階に分け、注視点 (Attention Point: AP) 解析 (以下 AP 解析) によって人間行動モデルを構築する手法を提案する (図 1)。

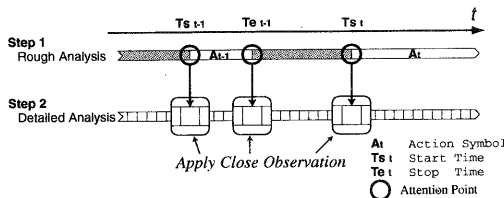


図 1: 注視点 (AP) を用いた 2 段階解析

注視点 (AP) とは、人間行動を解析する上で、一連の行動中で注意深い解析を必要とする部分を指す。これは、人間の行う単位動作 (つかむなど) の開始時点・終了時点や、ロボットが実際に獲得した行動を行ってみて失敗する個所などである。

第 1 段階では、ロボットは人間行動の大まかな解析を行い、粗い人間行動モデルを構築し、時間軸に沿って注視点 (AP) を抽出する。第 2 段階では、同じ人間行動に対して、第 1 段階で発見された注視点付近を重点的に解析する。複数のセンサを人

間行動の解析に利用できる場合には、センサの特徴に応じて第 1 段階と第 2 段階とでそれぞれ別のセンサを使い分けることもできる。このようにしてロボットが人間行動のうちの重要な部分だけを詳細に再解析することで、効率良く人間行動モデルを拡充していくことができる。

以降の章では、上記の AP 解析について詳しく説明し、データグローブと 9 眼ステレオビジョンを用いた人間行動モデル構築の実装法について説明する。最後に、本人間行動モデルの適用性を検証するために、実ロボットによる人間行動の認識・再現実験を行った結果について述べる。

## 2 人間行動の認識

本研究では、対象とするタスクをテーブル上における手作業に限定する。解析のための入力デバイスとして、データグローブ (CyberGlove & Polhemus 位置センサ) と 9 眼ステレオビジョンを用いた。ロボットは、データグローブからの入力データに基づいて粗い人間行動モデルを構築し、同時にそのモデルから注視点 (AP) を抽出する。第 2 段階では、ロボットは人間行動を再び観察するが、今回は注視点付近の行動のみに着目する。ここでは 9 眼ステレオビジョンを入力デバイスとして使い、注視点付近の距離データの解析を行う。本研究においては、ステレオデータ解析はデータグローブのデータ解析と比較して処理に時間がかかるため、注視点付近の距離データのみ解析を行うことで、効率的に人間行動モデルを拡充していくことが可能となる。

### 2.1 粗い人間行動モデル

ロボットは、人間の手作業における手の動きをデータグローブを使って解析することで、粗い人間行動モデルを構築する。

本研究では、人間の手作業を単位動作である「手動作」の連続体であると定義し、「手動作」を表 1 に示す属性を持つものと定義する。

解析には隠れマルコフモデル (HMM) によるジェスチャースポットティングを用い、人間の手作業全体が離散的に連続した “Action Symbol” (At) からなるものとして “Action Symbol” のセグメンテーションを行う。これについては、第 3 章で詳しく述べる。

このセグメンテーションされた “Action Symbol” を「手動作」として扱い、“Time Stamp (Ts, Te)” と “Hand” と “Position” の 3 つの属性を付与する

表 1: 手動作の属性

| Attributes    | Priority | value  |
|---------------|----------|--|
| Time Stamp    | 1(low)   | Absolute Time (start and stop time)                        |
| Action Symbol | 3(high)  | Power Grasp<br>Precision Grasp<br>Release, Pour, Hand Over |
| Hand          | 2        | Right, Left, Both  |
| Position      | 1        | Absolute Position in 3D space                              |
| Object Model  | 3        | Type of the Manipulated Object                             |

(表 1 参照).

この抽象化された「手動作」列を粗い人間行動モデルとする(図 2). このモデルを使うことで、人間行動時と全く同じ環境下においてはロボットは人間の手の動きと同等の動きを再現できるが、このモデルでは取り扱う対象物が何であるかという情報が欠落しているため、人間行動時と物体の配置が異なるようなより複雑な環境下においては、ロボットは簡単に作業を失敗してしまう。

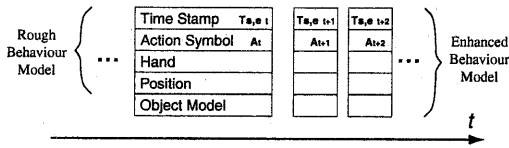


図 2: 人間行動モデル

## 2.2 注視点 (Attention Point: AP)

本研究では、注視点 (AP) としてそれぞれの「手動作」の開始時点と終了時点 ( $Ts_i, Te_i$ ) を選んだ。これは、この時点を境に人間はその行動を大きく変化させるため、この時点において物体の取り扱い方が大きく変わる可能性が高いためである。後からこの付近を視覚を用いて詳しく解析することで、ロボットはどの物体がどういう扱われ方をするかについての情報を獲得し、人間行動モデルに物体の情報を付加することができる。

例えば、人間が何かを “Power Grasp” したということが粗い人間行動モデルから分かったとする。しかし、この情報では人間がどこでどうつかんだかということは分かって、何をつかんだかは不明である。そこで、つかむ動作の開始時点を見視点

に設定し、人間の行動を再び観察して、視覚により注視点付近の画像中の手の存在する場所を解析することで、対象物がなんであるかの認識を行う。この場合は、つかむことで手により物体が隠蔽されることが分かり、またつかんだ時点での手の位置も分かっているため、AP 解析の時には、つかむ直前の画像を視覚によって獲得し、すぐ後に手が来るであろう場所を解析することで、つかまれた物体を認識することが可能になる。

## 2.3 注視点の詳細な解析

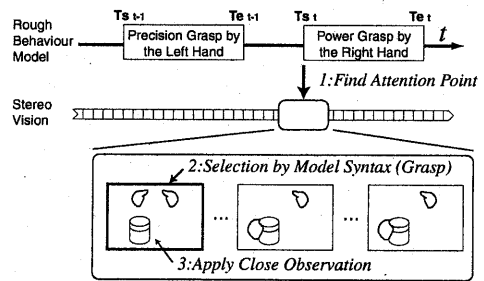


図 3: AP 解析例

本研究では AP 解析を行うための手段として、実際に 2 回人間の動作を観察する代わりに、一回の人間動作から粗い解析用と詳細な解析用の 2 種類の入力データを同時に獲得することを行った。ロボットは、人間の手作業をデータグループからのデータ (粗い解析用) によって解析する間、平行してステレオビジョン装置から得られる距離データ (詳細な解析用) を時系列に沿ってタイムスタンプと共に連続的に記録しておく。タイムスタンプは、人間行動モデルの “Time Stamp” と同期しているため、注視点 (AP) に対応するデータを解析しようとした場合、ロボットは対応する記録された距離データを簡単に獲得することができる (図 3)。

対応する距離データに対して、3次元テンプレートマッチング (3DTM) 法によって対象物体の認識を行う。認識手法の詳細については第 4 章で詳しく述べる。

## 2.4 優先度

「手動作」のそれぞれの属性は、優先度の項を持つ。優先度はタスクを遂行する上でのその属性の重

要度を示す。ロボットはまず人間行動モデルに記述されている通りにタスクを実行する。この時、人間行動時と物体の配置が異なるなどの理由でロボットの行動が失敗すると、ロボットは優先度の低い属性を無視し、再度同じタスクを遂行しようと試みる。この場合、“Action Symbol”など同じタスクを遂行する上で変化してはいけない項目は高い優先度を持ち、“Hand”や“Position”など、必ずしも人間動作時と同じである必要のない項目については優先度を低く割り振る。例を挙げると、「左手で場所Xにある物体Aをつかむ」という「手動作」をロボットが失敗したとすると、“Position”、“Hand”の順に対応する「手動作」から属性を取り去り、この場合最終的には「物体Aをつかむ」という「手動作」に抽象化してもう1度動作を試みる。このように、タスクの遂行を中断する前に制約条件を少なくすることで、タスクの継続性を高めている。

AP解析によって構築された人間行動モデルは高度に抽象化されており、また優先度項により実行時に環境に応じて抽象度の度合いを変化させられるため、例えば物体の配置や数が異なる環境における適用性が高い。第5章でこの例を示す。

### 3 ジェスチャースポットティングによる粗い人間行動モデルの構築

粗い人間行動モデルにおける“Action Symbol”を求めるために、人間がなんらかの手作業を行っている最中に、隠れマルコフモデル(HMM)を利用したジェスチャースポットティングによって一連の行動の中に離散的に現れる各“Action Symbol”の切り出しを行った。本研究では、“Action Symbol”として次の6つの動作をジェスチャー(表1のAction Symbolに示す5つ+トレーニング用のOK sign)としてモデル化した。両手分のデータグループからの入力データは左右独立にスポットティング処理され、両手を用いるジェスチャーは左右のスポットティング認識の結果を合わせることで認識し、“Action Symbol”の抽出を行っている。

この章ではHMMの概要、スポットティング認識、そして両手のデータグループ(CyberGlove)を用いた実装レベルでのシステムの説明を行う。

表 2: ジェスチャーの定義

| Gesture         | Primitives  | Action                                    |
|-----------------|-------------|---|
| Power Grasp     | cls+sp      | Power-grasp from open position            |
| Precision Grasp | prc+sp      | Precision-grasp from open position        |
| Pour            | cls+roll+sp | Power-grasp, and roll the wrist           |
| Hand-over       | prc+forw+sp | Precision-grasp, move forward, and back   |
| Release         | opn+sp      | Open a grasp hand                         |
| OK-sign         | ok+sp       | Make a circle with thumb and index finger |
| Garbage         | gb          | A filler model for spotting               |
| Start,End       | sil         | Silence at the start & end                |

#### 3.1 HMM

HMMは不確定な時系列のデータをモデル化するために有効な統計的手法であり、音声認識における重要な基礎技術の一つとなっている[4]。HMMの特徴は、単純マルコフモデルの状態間の遷移と状態における出力がどちらも確率的であり、状態が出力系列から一意に決まらないところにある。これが「隠れ」マルコフモデルと呼ばれる所以である。出力系列に含まれるノイズを確率的に処理出来る土、モデルのシンボルと時系列データを結びつける各状態の理論的展開が容易な事から、音声のみならずジェスチャー認識にも使われる事が多い[5]。

#### 3.2 ジェスチャースポットティング

ジェスチャーのスポットティング認識とは、連続した動作の中から時間的位置が未知のジェスチャーの判別と時系列上のセグメンテーションを同時に行うものである。これにより、ジェスチャーを動作の開始と終了を意識することなく認識する事が可能となる。

標準的なHMMによるスポットティング認識では、キーワードとなるHMMとFiller(Garbage)HMMを並列に繋げ、対象外の動きをFiller HMM内の状態に落とし込む方法がとられる[6]。音声の場合、Filler HMMを全音素のHMMから用意する事が可能だが、ジェスチャーの場合明確な要素となる動作の定義は無い。そのため、アドホックに作成するか、閾値を出すための閾値モデルを作るなどの工夫が必要となる。本研究ではスポットティング対象の語彙数が少ないため、Filler HMMを独自に作成した。

### 3.3 データグローブを用いた認識

本研究では左右のデータグローブ (CyberGlove) と三次元位置センサ (Polhemus) を入力として, HMM を用いてスポッティング認識を行った. なお, システムの一部は, Hidden Markov Model Toolkit (HTK) [7] に基づいている.

出力データとして, 片手につき 48 次元の特徴点を使用している. CyberGlove から得られる 18 次元の手指の曲げ  $\{r_1 \dots r_{18}\}_t$ , Polhemus から得られる 6 次元の手の姿勢  $P_t = \{x, y, z, \alpha, \beta, \gamma\}_t$ , それにおおの速度情報を加えて, 計 48 次元としている. Polhemus からの位置情報  $P_t$  は, 作業者の向きや位置に依存しない認識を可能とするため, 常に 1 フレーム前の状態  $P_{t-1}$  を座標の起点とするフレーム間の速度  ${}^{t-1}P_t$  に変換してから特徴点として使用している (図 4).

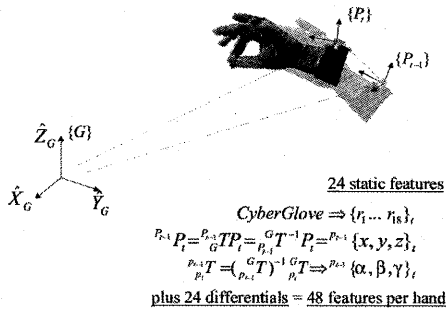


図 4: 特徴ベクトルの定義

ジェスチャーの定義では各々に独立した HMM を割り当てず, 構成要素の HMM を定義し, それらを繋げたものを使用した (表 2). 各ジェスチャーが要素を共用する事により, 学習データの不足を補い, 学習効率を向上させる事が可能となっている. *cls*, *tsu*, *roll*, *mae*, *opn*, *ok* は 5-state left-right (初めから終わりまで遷移が一方通行) HMM でモデル化されている. *sil* はトレーニングの際の開始&終了状態, *sp* はジェスチャー終了時の短い静止状態, *gb* はジェスチャー間の動きを他の要素よりも高い確率で拾うガベージコレクタとして使用する.

### 3.4 解析結果

図 5 に, 本ジェスチャースポッティングで用いた HMM 上でのジェスチャーの状態遷移を示す. 本シ

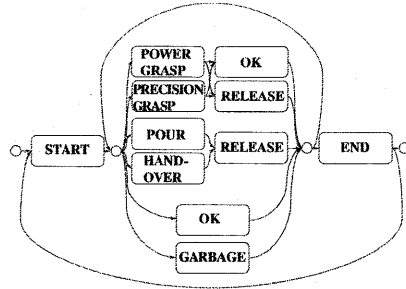


図 5: ジェスチャーの状態遷移

表 3: ジェスチャーの認識結果

|            | Left     | Right    |
|------------|----------|----------|
| % Accuracy | 98.89%   | 95.56%   |
| N,D,S,I    | 90,0,0,1 | 90,0,4,0 |

% Accuracy =  $\frac{N-D-S-I}{N} \times 100\%$   
 (N)umber of gestures, (D)eleation error,  
 (S)ubstitution error, (I)nsertion error

ステムでは, 左右のデータグローブそれぞれについて 30Hz のサンプリング周期でデータを取ることができ, サンプリング周期に同期して時間遅れ無くジェスチャースポッティングを行うことができる.

モデルの学習は片手ずつ行われ, 右腕は文法にそったトランスクリプト付き学習データを 4 パターン (並び方) 各 5 回分とガベージモデル用データを 10 回分, 計 184 秒 (5520 フレーム, 内ガベージ用は 1670 フレーム) 分を使用した. 左腕についても同様である. ジェスチャーの単語認識は学習データとは別パターン (並び方) のテストセットで実験を行った. テストセットは 3 パターンのジェスチャー配列を 5 回ずつ, 計 109 秒分を使用した. 結果を表 3 に示す.

## 4 ステレオビジョンを用いた人間行動解析

本章では, 人間行動モデルに物体の情報を付加するために, ステレオビジョンと 3 次元テンプレートマッチング法によって人間が扱った物体の種類を推定する方法について説明する. 本研究では, コマツの製品である 9 眼のマルチベースラインステレオシステム [8] を高周期で距離データを生成する目的で使用した. カメラ配置については, 図 6 に示す

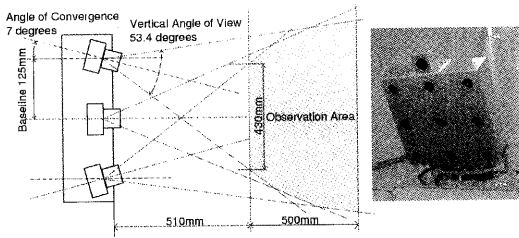


図 6: ステレオビジョンシステム

ように、ベースラインを大きく取り外周のカメラを内側に傾けることで、近距離の物体を見るロボットの視覚として適した形に設計した。測定範囲は可変であり、今回は最も近距離が測定可能なように測定範囲を設定した (510mm - 1010mm)。

以下、物体の認識に用いた 3 次元テンプレートマッチング (3DTM) 法について述べ、ステレオビジョンと 3DTM の組み合わせによる人間行動の詳細な解析方法について述べる。

#### 4.1 3次元テンプレートマッチング

3次元テンプレートマッチング (3D Template Matching: 3DTM) [9] とは、3次元距離データ中において、あらかじめ対象のおおまかな位置と 3次元幾何モデルが既知であるという前提の下で、対象物の正確な位置と姿勢を推定する手法である。

3DTM では、対象物体の 3次元モデルを、レンジデータなどによって得られた対象物体の 3次元データ中に投影し、モデルの頂点とそれに最も近い周囲の 3次元点との重み付き距離の和の平均を求める。この和を指標として、これが最小となる状態に収束するようにモデルの位置・姿勢を変化させていく。収束地点でのモデルの位置・姿勢が推定値となる。

#### 4.2 3DTM を用いた人間行動解析

ロボットは、次の手順で注視点における距離データの解析を行う。

1. 対象タスクはテーブル上での手作業であるという前提があるため、得られた距離データから平面フィッティングにより背景を抜き取り、その上にある物体をセグメンテーションする。

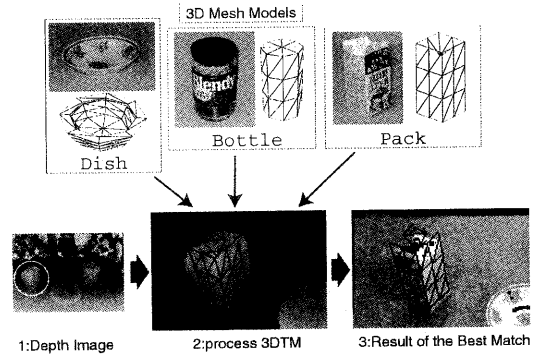


図 7: 3DTM による物体認識

2. 物体は安定姿勢にあるとの仮定をおくと、セグメンテーションにより物体のおおまかな位置・姿勢が分かるため、それを初期位置とし、セグメンテーションされた物体について手持ちの全ての 3次元モデルを使用して 3DTM をかける (図 7)。
3. それぞれの物体について、指標である重み付き距離の和の平均が最も小さくなるモデルを対象物のモデルであると推定する。

3DTM は投影されたモデルの初期位置に敏感であり、また対象物体が隠蔽されていない程良い結果をもたらす。人間行動モデルから、“Action Symbol” 属性と “Time Stamp” 属性より手によって隠蔽されていない物体を解析するにはどの時点の距離データを解析すればよいか分かり、“Position” 属性によって距離データ中の大体どの位置に物体があるのかが分かる。これと上記の物体のセグメンテーション結果を付き合わせて 3DTM をかけることで、精度高く対象物体の認識を行うことが可能になる。

表 4: 3DTM の結果: M推定量 (ローレンツ)

| Objects<br>(in Depth Data) | Models      |             |             |
|----------------------------|-------------|-------------|-------------|
|                            | Pack        | Dish        | Bottle      |
| Pack                       | <u>0.25</u> | 1.30        | 0.55        |
| Dish                       | 2.08        | <u>0.65</u> | 1.43        |
| Bottle                     | 0.92        | 1.20        | <u>0.37</u> |

表 4 に、今回の実験で用いた物体に対して 3DTM をかけた結果を示す。表中の数字は最小二乗法の一般形である M 推定量による重み付き距離を正規化

した値の和の平均を示し、下線の引いてある数字は最も高い一致度（小さい方が一致度が高い）を表す。この結果から、物体の認識が正確に行われていることが分かる。

## 5 ロボットによる認識と再現

### 5.1 プラットフォーム

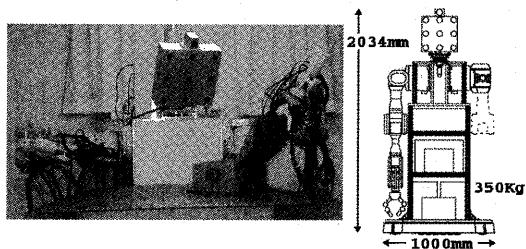


図 8: 実験プラットフォーム

ロボットによる人間行動の学習及び人間・ロボット協調作業のための実験プラットフォームとして、図 8 に示すロボットを開発した。我々の研究では、人間行動のうちでも人間の手作業に着目しており、ロボットによる人間の手作業の獲得およびその実行を実現することが目標である。そのための実験プラットフォームには、人間の上半身と同様の機能が求められ、特に人間と同等の眼の機能と手の機能及び上半身の機能を持たせることが重要になる。本ロボットの特徴は次の通りある。

- 9 眼ステレオビジョンにより環境の 3 次元認識が可能である（眼の機能）。
- 2 台の 7 自由度ロボットアームを双腕として持つ。右腕には 4 本指のハンドが、左腕には 3 本指のハンドがついている。それぞれの指は 3 自由度であり、指先にはフォーストルクセンサが装着されている（腕の機能）。
- 視点や腕を自由に動かすために、ロボット全体が 2 次元平面上を自由に動くことができる（上半身の機能）。
- CORBA [10] を基盤にしたソフトウェアアーキテクチャによって、データグローブなどの外部デバイスが容易に接続可能である。

### 5.2 実ロボットによる認識・再現実験

AP 解析によって構築した本人間行動モデルの有効性を検証するために、次に示すロボットによる人間行動の認識・再現実験を行った。この実験では、人間はロボットの前で容器 A をつかみ、次に容器 B をつかんでその中身を容器 A の中に注ぐ。ロボットはその作業を観察しており、人間行動モデルを構築する。次に、その行動モデルに従って、ロボットはその動作の再現を行う。再現時には、モデルの優先度の効果を確認するために、人間動作時とは若干異なる環境で再現動作を行い、モデルの適用性を検証した。

#### 5.2.1 人間行動の認識

解析の第 1 段階として、ロボットはデータグローブからの入力を用いて人間の作業の観察を行った。この時、ジェスチャースポッティングによって、ロボットは粗い人間行動モデルを構築した。同時に、つかむ・注ぐ動作の開始点を注視点 (AP) として抽出した（図 9 の上段）。この解析によって、モデルに手の動作の情報が付加された。

解析の第 2 段階として、第 1 段階において平行して記録しておいた距離データ群の中から、注視点 (AP) 付近の距離データをタイムスタンプを指標に読み込み、3DTM による対象物体の認識を行った（図 9 の中段）。この解析によって、モデルに物体の情報が付加された。

#### 5.2.2 人間行動の再現

次にロボットは、構築された人間行動モデルに従って、同じ作業の再現を行った。

モデルの異なる環境における適用性を検証するために、人間作業時には存在しなかった新しい物体（皿）をテーブル上に追加し、また物体を人間作業時とは異なるように配置し直した。再現時にテーブル上の物体を認識する手段として、第 4 章で説明した方法と同じ方法を用いた。

実験の結果は、ロボットが正しい物体を選択し作業を遂行できたことを示している（図 9 の下段）。ロボットの再現した動作を検証すると、次の 2 点において優先度の効果が現れていることが分かった。

- 対象物体がモデルで記述されている位置に無い場合には、ロボットは優先度の低い“Position”項を省略し、視界全体から対象物体を探して

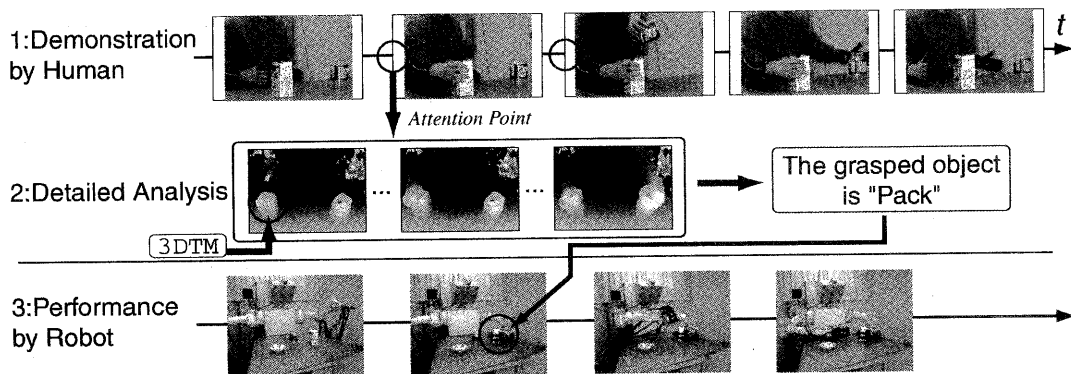


図 9: 認識・再現実験

いる。これは、特に似た形状の物体が複数ある場合には、正しい物体を効率よく探すために有効である。

- 物体の配置が変わったため扱うべき物体が本来扱うべき腕の操作範囲の外にある場合には、ロボットは優先度の低い“Hand”項が省略し、他の届く方の腕でつかみにいっている。

優先度項により、「手動作」毎に抽象度を変更し、作業の遂行を優先した上でなるべくモデルに忠実に作業を再現することができた。

## 6 まとめ

本論文では、データグローブと9眼ステレオビジョンを用いたAP解析により、人間行動モデルを構築する手法について述べた。またこのモデルは、シンボル化と優先度項によって人間行動時とは異なる環境における適用性を高めており、実ロボットによる人間作業の認識・再現実験を行うことによって、その適用性の検証を行った。

## 参考文献

- [1] K. Ikeuchi and T. Suehiro, “Toward an Assembly Plan from Observation Part I: Task Recognition With Polyhedral Objects,” *IEEE Trans. Robotics and Automation*, 10(3):368–384, 1994.
- [2] Y. Kuniyoshi, et al., “Learning by watching,” *IEEE Trans. Robotics and Automation*, 10(6):799–822, 1994.
- [3] H. Kimura and T. Horiuchi and K. Ikeuchi, “Task-Model Based Human Robot Cooperation Using Vision,” *IROS*, 2:701–706, 1999.
- [4] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected, Applications in Speech Recognition,” *In Proc. of the IEEE*, vol. 77, no. 2, pp.257–285, February 1989.
- [5] T. Starner and A. Pentland, “Real-time American Sign Language recognition from video,” *IEEE International Symposium on Computer Vision*, Coral Gables, FL. 265–270, 1995.
- [6] K. M. Knill and S. J. Young, “Speaker Dependent Keyword Spotting for Accessing Stored Speech,” *Cambridge University Engineering Dept., Tech. Report*, No. CUED/F-INFENT/TR 193, 1994.
- [7] S. J. Young, “Hidden Markov Model Toolkit V2.2.,” Entropic Research Lab Inc., Washington DC, January 1999.
- [8] <http://www.komatsu.co.jp/research/study56.htm>
- [9] M. D. Wheeler and K. Ikeuchi, “Sensor Modeling, Probabilistic Hypothesis Generation, and Robust Localization for Object Recognition,” *IEEE Trans. PAMI*, 17(3):252–265, 1995.
- [10] Common Object Request Broker Architecture, OMG, July, 1995.