

# 遮蔽近傍時刻を用いた見えない作用推定に基づく 道具利用タスク学習法

## Invisible Functions' Estimator Based on Temporal Sequence of Visual Hierarchies and Motion Features

長濱 虎太郎 (東京大学)

三宅 健司 (東京大学)

山崎 公俊 (信州大学)

矢口 裕明 (東京大学)

岡田 慧 (東京大学)

稲葉 雅幸 (東京大学)

Kotaro NAGAHAMA\*, Kenji MIYAKE\*, Kimitoshi YAMAZAKI†,  
Hiroaki YAGUCHI\*, Kei OKADA\*, and Masayuki INABA\*

(\*The University of Tokyo, †Shinshu University)

The purpose of this research is to estimate functions of objects during the observation of tasks to move an object using another object. When an object is loaded onto another object or put into another object, one or more objects often becomes unable to be recognized mainly because of the occlusion. We propose a novel method called “V-FES” to estimate the states of unrecognized objects using the time-series data of motion features and visual hierarchies of them. We tested the method using sparse tracking results. These results suggested the effectiveness of our approach.

### 1 緒言

本研究の目的は、生活支援ロボットが人の道具を利用した日常タスクを視覚学習し再現するためのシステムの構築である。タスクの視覚学習法は従来より研究されているが [1, 2], 多くの場合、観察対象の追跡には常に成功しているとして、作用のシンボル化等に主眼が置かれてきた。

本稿で取り上げるタスクは、箱に物をしまう・トレーで物を運ぶ等の、道具に物を重ねる・入れる、そしてその状態で物を運ぶようなタスクである。つまり、1) 何に何を重ねるか、入れるか、そして 2) それらをどこへ持っていくかという、1) 対象の道具的利用法の推定、2) タスク目的の推定をおこなわせたい。このようなタスクでは、例えば物に物を重ねる場合や入れる場合に、観察対象がカメラに映らない状況が発生する。また、観察中に対象がフレームアウトする状況も容易に起こり得ることから、どのような追跡手法を用いても、観察対象を正しく認識する、つまり位置を正しく推定することは難しいと言える。

一般に観察対象が認識できなくなる理由としては、A1)

センサの観測領域に含まれない場合 (観測不可能), A2) センサの観測領域に含まれているが、他の物体に遮蔽されている場合 (観測不可能), A3) 観測可能だが、物体を特定するのに十分な視覚特徴がない場合 (小さすぎる、見た目の区別がつかない等) が考えられる。また、認識された場合でも、認識結果が誤っている場合として B1) 本当は観測不可能であるのに、誤認識してしまう場合、B2) 観測可能だが、誤った場所に認識してしまう場合、が考えられる。観察学習の際には、対象の道具的利用によって A2) が起こる。また観察者のセンサの動きが追いつかない場合に、フレームアウトによって A1) が、対象との距離が遠くなることで A3) が容易に起こり得る。

遮蔽下にある対象の認識法としては、対象のテンプレートを変形しながら追跡することで、部分的な遮蔽に対応する手法 [3, 4], ある程度の遮蔽に対応した特徴量を用いた探索手法 [7] 等が研究されている。しかしこれらの手法では、対象の道具的利用で発生する、A1) や A2) の完全な遮蔽には対応できない。遮蔽下では周辺の視覚特徴を記憶して追跡する手法 [5, 6] も研究されており、完全な遮蔽下でも追跡を可能としているが、完全な遮蔽下での大きな移動やフレームアウトが起こると、追跡は不可能だと言える。

本研究では A1) から A3) へ対応するため、大幅な遮蔽の後でも対象の位置を特定できるように、SIFT [7] や LINE-MOD [8] 等の探索手法の認識結果を入力として用いる。さらに B1) と B2) に対応するため、探索手法の検出の評価を厳しくし、偽陽性のない認識結果を期待する。これにより増加する A3) のミスについては、認識されない時区間を埋めるための知識を用いて対応する。そして探索成功時刻のみの認識結果を用い、各時刻の対象の遮蔽下・非遮蔽下での状態推定をおこない、いつ何の作用がおこなわれているか、タスク目的、そして対象の道具的機能を推定するための機構、V-FES (Visual Hierarchy-based Function ESTimator) を提案し評価する。

### 2 視覚重畳関係を利用する V-FES の構成

本稿で提案する V-FES を用いる観察学習システムの構成を Fig.1 に示す。V-FES の入力は、観察対象の探索に成功し

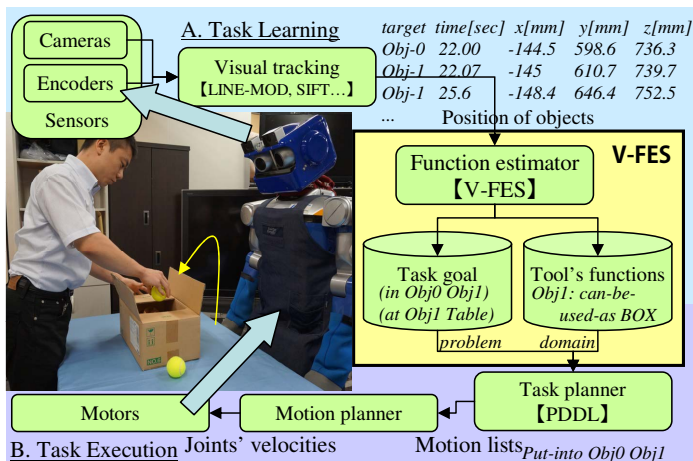


Fig. 1: Task learning system with V-FES

た時刻のみの認識結果（位置，Fig.1，右上）である．V-FESでは 1) 対象の道具的利用法，2) タスク目的を推定し，その結果はタスクプランナへ送られる．タスクプランナでは STRIPS[10] タイプのオペレータを用いてシンボリックなタスク再現計画 [1] をおこなうことで，ロボットの再現のための動作列を決定するものとする．

V-FES で推定し，タスクプランニングで利用する 1) 対象の道具的利用法は，a) 他の対象を載せることのできる“MAT”機能，b) 載せた状態で動かすことのできる“TRAY”機能，c) 他の対象を入れることのできる“STORAGE”機能，b) 入れた状態で動かすことのできる“CONTAINER”機能の四種類とする．2) タスク目的としては，a) 対象をどこへ移動するか，b) 対象同士の関係をどのようにするか（載せる，入れる），を計算する．これにより，道具を用いた片付けタスクの推定・記憶をおこなう．

V-FESの各時刻における対象の作用推定には，まず二つの対象の 1) 視覚重畳関係と 2) 運動特徴というメタ情報を推定し，メタ情報から作用を推定する手法 [4] を用いるものとする．この手法では，その接触・近接関係が無いと作業に失敗するような重要な面同士の関係 [11] を，接触・近接面から十分離れた場所から観測することで，遮蔽関係を観測し利用するものである．しかし [4,11] では，各時刻で独立にメタ情報を推定し，組み合わせて作用を推定していたため，対象が全く認識されない時刻には対応できていなかった．そこで V-FES では，各メタ情報を前後の時刻の情報も利用して推定する．Fig.2 に，V-FES の処理の流れを示す．メタ情報の時系列データを，五つの知識（KNO-1 から KNO-5）に基づいたアルゴリズムを用いて，順に修正していく．最終的な時系列メタ情報から各時刻でおこなわれている作用を推定し，前述の道具機能とタスク目的を算出するものとする．次章では，メタ情報の時系列データを修正していくためのアルゴリズムについて，順に説明する．

### 3 時間近傍特徴を用いた見えない特徴推定

#### 3.1 作用推定のためのメタ情報

本研究で，二つの対象間の作用の推定をおこなうための手がかりとする，メタ情報は以下の通りとする．

- 対象の視認性
  1. 認識の可否 Visibility ( $s_V^k$ )
 

対象  $k$  が観測可能であるだけでなく，認識器により検出がなされており，位置が特定できている場合を“Recognized” (“Rec.”)，それ以外を“Unrecognized” (“Unrec.”) とする．
  - 対象の運動特徴
    2. 各対象の動きの有無 Movements ( $s_M^k$ )
 

各対象ごとに存在する状態（小文字の  $s$  で表す）で，対象が動いているときは“Yes”，動きが少ない場合は“No”とする．動きが分からない場合は“Unknown”とする．
    3. 対象同士の距離 Distance ( $S_D$ )
 

対象同士の距離が近い場合は“Near”，遠い場合は“Far”とする．距離が分からない場合は“Unknown”とする．
    4. 対象同士の運動随伴性 Co-movements ( $S_C$ )
 

対象がもう一方の対象を動かしている作用の推定に用いられるメタ情報である．対象が動いている際に，対象同士の運動ベクトルが類似している場合を“Yes”，そうでない場合を“No”とする．分からない場合は“Unknown”とする．
    - 対象同士の視覚重畳関係
      5. 対象同士の視覚重畳関係 Visual Hierarchy ( $S_H$ )
 

対象同士の重なり方を表すメタ情報であるが，タスクに重要な接触・近接面から十分離れた場所から観測し，かつ近接関係を確認することで，物体を重ねる・入れるという真の物体の重なり方と近似できると考える．重なり関係が
















1. Visibility		3. Distance		5. Visual Hierarchy	
a) Unrecognized		a) Unknown		a) Unknown	
b) Recognized		b) Far		b) Independent	
		c) Near		c) One-way (Obj-0)	
2. Movements		4. Co-movements		d) One-way (Obj-1)	
a) Unknown		a) Unknown			
b) No		b) Yes			
c) Yes		c) No			

Fig. 3: Types and color legends for V-FES

存在しない場合を“Independent” (“Indp.”)，一方が全面にある場合（片側遮蔽）を“One-way”，特に対象  $k$  が前面に存在する場合を“One-way (OBJ- $k$ )”とする．

以後，これらのメタ情報の推移は，Fig.3 に示すような凡例色を用いて表現するものとする．

#### 3.2 追跡結果からの運動情報推定

まず，認識の成否  $s_V^k$  を決定する．一定時間  $t_C$  毎に時区間 ( $T_{All} \in T_0, T_1, \dots, T_{fin}$ ) を設定し，各時区間に対象  $k$  の認識結果が存在すれば，その区間で  $s_V^k(T_i)$  は“Rec.”，そうでなければ“Unrec.”とする．対象  $k$  ( $k = 0, 1$ ) が時区間  $T_i$  で“Rec.”であった場合，認識結果の三次元位置  $x^k(t)$  [mm] ( $t_C \cdot i \leq t < t_C(i+1)$ ) を平均してその時区間での位置  $\bar{x}^k(T_i)$  とする．対象が一度も検出・認識されなかった時区間では，平均位置は計算しない．以後，時区間  $T_i \in T_{All}$  を単に時刻と呼び，時刻集合を時区間と呼ぶものとする．

次に，各時区間における各対象の運動の有無を計算する．時刻  $T_i$  の対象  $k$  の速度  $v^k(T_i)$  (mm/ $t_C$ [sec]) は，

$$v^k(T_i) = \bar{x}^k(T_i) - \bar{x}^k(T_{i-1}) \quad (1)$$

として求まる．この速度について閾値計算をすることで，運動の有無  $s_M^k(T_i)$  を評価する．ただし， $\bar{x}^k(T_{i-1})$  と  $\bar{x}^k(T_i)$  が計算されていない場合，つまり  $s_V^k =$  “Unrec.” の場合は“Unknown”とする．

$$s_M^k(T_i) = \begin{cases} \text{“Unknown”}, & \text{if } s_V^k(T_{i-1}) = \text{“Unrec.”} \\ & \text{or } s_V^k(T_i) = \text{“Unrec.”} \\ \text{“Yes”}, & \text{if } \|v^k(T_i)\| \geq v_{movep} \\ \text{“No”}, & \text{otherwise} \end{cases} \quad (2)$$

ただし  $v_{movep}$  は，運動の有無を評価するための定数とする．

続いて，観察対象同士の距離の推移を計算する．時刻  $T_i$  において，二つの対象が各々の位置  $\bar{x}^0(T_i)$ ， $\bar{x}^1(T_i)$  に認識されているとき，距離の評価値  $S_D(T_i)$  は以下のように計算される．

$$S_D(T_i) = \begin{cases} \text{“Unknown”}, & \text{if } s^0(T_i) = \text{“Unrec.”} \\ & \text{or } s^1(T_i) = \text{“Unrec.”} \\ \text{“Near”}, & \text{if } \|\bar{x}^0(T_i) - \bar{x}^1(T_i)\| < D_{nearp} \\ \text{“Far”}, & \text{otherwise} \end{cases} \quad (3)$$

ただし  $D_{nearp}$  は，距離の評価のための正の定数とする．

さらに，観察対象同士の視覚重畳関係を評価する．常に対象領域が一部でも認識可能であれば，重なり領域を推定し，画像と対象の表面のテクスチャを比較することで，上下関係を推定することが可能である [4]．しかし本稿では対象が認識できない状況，対象がカメラに映らない状況をも扱う．そこで，Fig.1 右上に示すような対象の三次元軌道と，各時刻での認識の可否 ( $s_V^k(T_i)$ ) のみから，視覚重畳関係を推定する手法を検討する．二つの対象がいずれも認識されており，かつ十分に離れている場合には，2つの視

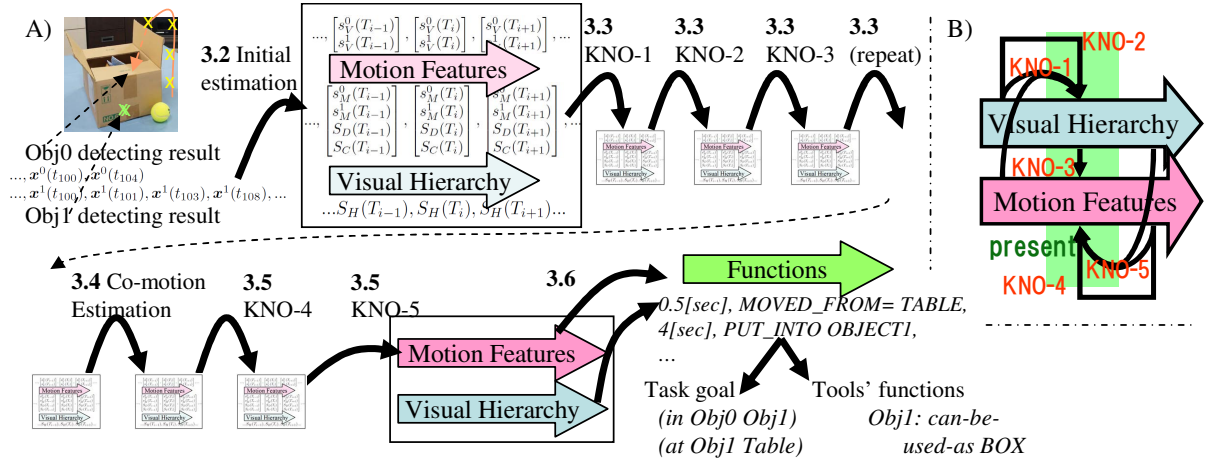


Fig. 2: A) System flow of V-FES. B) Concept to fix meta-information in V-FES

覚重畳関係は“Indp.”と決定できる．

$$S_H(T_i) = \begin{cases} \text{“Indp.”, if } s_V^0(T_i) = s_V^1(T_i) = \text{“Yes”} \\ \quad \text{and } S_D(T_i) = \text{“Far”} \\ \text{“Unknown”, otherwise} \end{cases} \quad (4)$$

この段階で決定できるメタ情報は以上である．

### 3.3 知識 KNO-1 から KNO-3 の利用

1つ目の知識 KNO-1 は，過去の運動特徴と過去の視覚重畳関係から，現在の視覚重畳関係を推定するための知識である．“KNO-1: 対象同士が近づいた後に一方の対象しか見えなくなったら，その後にもう一方が見えるまでは片側遮蔽関係だろう”とする．KNO-1 が成り立つための前提条件は，対象が重要な見え隠れをする近くの時刻で，必ず対象が認識可能であることである．観測者から対象が観測可能であることが必要となることから，教示者は対象同士に重要な遮蔽関係が発生していないときに観測者へ対象を見せるようにすれば良い．

KNO-1 を適用するアルゴリズムは，次のようである．ただし，各時区間  $\mathbb{T}$  の始めの時刻と終わりの時刻を，それぞれ  $T_{init}(\mathbb{T})$ ,  $T_{fin}(\mathbb{T})$  とし， $T_{i+1} = T_i + \Delta T$  のように書くものとする．

1. 書き換え対象時区間  $\mathbb{T}_{K1,i}$  を探索する．ただし

$$\begin{aligned} \mathbb{T}_{K1,i} &\subset \mathbb{T}_{All}, S_D(T_{init}(\mathbb{T}_{K1,i}) - \Delta T) = \text{“Near”}, \\ s_V^0(T_{init}(\mathbb{T}_{K1,i})) &= \dots = s_V^0(T_{fin}(\mathbb{T}_{K1,i})) = \text{“Rec.”}, \\ s_V^1(T_{init}(\mathbb{T}_{K1,i})) &= \dots = s_V^1(T_{fin}(\mathbb{T}_{K1,i})) = \text{“Unrec.”}. \end{aligned}$$

2. 各  $\mathbb{T}_{K1,i}$  で， $S_H(T_n) = \text{“Unknown”}$  である  $T_n \in \mathbb{T}_{K1,i}$  があれば， $S_H(T_n) = \text{“One-way (Obj-0)”}$  と書き換える．
3. 対象 0 と対象 1 を入れ替えて，1., 2. をおこなう．

2つ目の知識 KNO-2 は，過去の視覚重畳関係から，現在の視覚重畳関係を推定するための知識である．“KNO-2: 片側遮蔽の後に観察対象が両方見えなくなったら，手前の観察対象の認識失敗だろう”とする．KNO-2 が成立するための前提条件は，タスクに重要な相互作用をしよう対象は，全てロボットが検出可能であることと，対象が重要な見え隠れをする近くの時刻で，必ず対象が認識可能であることである．単なる観察対象の認識失敗であれば観察対象同士の視覚重畳関係は変化しないとすれば，KNO-2 は次のように利用できる．

1. 書き換え対象の時刻  $T_{K2,i} \in \mathbb{T}_{K2}$  を探索する．ただし， $\mathbb{T}_{K2} = \{T_{K2,i} \mid T_{K2,i} \subset \mathbb{T}_{All}, S_H(T_{K2,i} - \Delta T) = \text{“One-way”}, s_V^0(T_{K2,i}) = s_V^1(T_{K2,i}) = \text{“Unrec.”}\}$
2.  $T_{K2,i} \in \mathbb{T}_{K2}$  について， $S_H(T_{K2,i}) = S_H(T_{K2,i} - \Delta T)$  として書き換える．

3つ目の知識 KNO-3 は，視覚重畳関係から，運動特徴のうち近接関係を推定するための知識である．“KNO-3: 近接した後に片側遮蔽関係が発生しているということは，物体が近いということ”とする．KNO-3 が成立するための前提条件は，二つの対象と観測点が一直線上に並んだ状態で，対象同士の距離が有意に変化しないことである．観測点と2つの対象の位置関係に関する制約であることから，教示者が観察者であるロボットへ，対象が極端に見えにくならないように注意してやれば良い．

KNO-3 を適用するアルゴリズムは，次のようである．

1. 書き換え対象の時刻  $T_{K3,i} \in \mathbb{T}_{K3}$  を探索する．

$$\begin{aligned} \mathbb{T}_{K3} &= \{T_{K3,i} \mid T_{K3,i} \subset \mathbb{T}_{All}, \\ S_D(T_{K3,i} - \Delta T) &= \text{“Near”}, S_H(T_{K3,i}) = \text{“One-way”}\} \end{aligned}$$

2. 各  $T_{K3,i} \in \mathbb{T}_{K3}$  で， $S_D(T_{K3,i}) = \text{“Unknown”}$  であれば， $S_D(T_{K3,i}) = \text{“Near”}$  と書き換える．

知識 KNO-1 と KNO-2 は過去の状態を利用するが，KNO-3 は現在の状態を利用するため，書き換えは一度では終了しない．書き換えが完全に終了するまで，KNO-1 から KNO-3 までの知識を繰り返し利用するものとする．

### 3.4 運動随伴性の評価法

続いて，二物体の運動随伴性の評価をおこなう．いずれの物体も速度が計算できる場合には，二物体の速度を用いて運動随伴性を評価する [4]．時刻  $T_i$  での運動の随伴性評価  $S_C(T_i)$  は，以下のように求められる．

$$S_C(T_i) = \begin{cases} \text{“Unknown”}, & \text{if } -h_M(T_i, 0) \text{ and } -h_M(T_i, 1) \\ \text{“No”}, & \text{if } (h_M(T_i, 0) \text{ and } h_M(T_i, 1) \text{ and } \neg h_c(T_i)) \\ & \text{or } (h_M(T_i, 0) \text{ and } \neg h_M(T_i, 1)) \\ & \text{or } (\neg h_M(T_i, 0) \text{ and } h_M(T_i, 1)) \\ \text{“Yes”}, & \text{otherwise} \end{cases}$$

$$h_M(t, k) \triangleq \|\mathbf{v}^k(t)\| \geq v_{thr}, h_c(t) \triangleq \|\mathbf{v}^0(t) - \mathbf{v}^1(t)\| \geq E_a$$

ただし  $v_{thr}$  と  $E_a$  はそれぞれ、運動の有無と随伴性の有無を評価するための正の定数とする。しかしこの評価は、物体の速度が計算できている場合のみ可能である。したがって物体が速度が計算されていない場合、つまり  $s_v^k(T_i) = \text{“Unrec.”}$  あるいは  $s_v^k(T_{i-1}) = \text{“Unrec.”}$  の場合には、評価できない。この場合に対応する知識が KNO-5 であるが、この段階では計算を保留し、 $S_C(T_i) = \text{“Unknown”}$  とするものとする。KNO-5 の詳細については、続いて述べる。

### 3.5 知識 KNO-4 と KNO-5 の利用

4 つ目の知識 KNO-4 は、未来の運動特徴から、現在の運動特徴を推定するための知識である。“KNO-4: 他の対象に隠されていない対象を見失ってから再度発見した場合には、単なる認識失敗である。位置が変わっていないければ「止まっている」、そうでなければ「動いている」とする。KNO-4 が成立するための前提条件は、対象が見えていない間に、タスクに関連する重要な動作（停止含む）がおこなわれていないことである。

KNO-4 を適用するアルゴリズムは、次のようである。

1. 書き換え対象の時区間  $\mathbb{T}_{K4,i}$  を探索する。ただし

$$\begin{aligned} \mathbb{T}_{K4,i} &\subset \mathbb{T}_{All}, \\ s_V^0(T_{init}(\mathbb{T}_{K4,i})) &= \dots = s_V^0(T_{fin}(\mathbb{T}_{K4,i})) = \text{“Unrec.”}, \\ S_H(T_{init}(\mathbb{T}_{K4,i})), \dots, S_H(T_{fin}(\mathbb{T}_{K4,i})) \\ &\neq \text{“One-way (Obj-0)”}. \end{aligned}$$

2. 各時区間  $\mathbb{T}_{K4,i}$  において、 $\|\bar{x}^0(T_{init}(\mathbb{T}_{K4,i})) - \bar{x}^0(T_{fin}(\mathbb{T}_{K4,i}))\| > A$  であれば、 $T_n \in \mathbb{T}_{K4,i}$  について  $s_M^0(T_n) = \text{“Yes”}$ 、そうでなければ  $s_M^0(T_n) = \text{“No”}$  とする。ただし、 $A$  は正の定数とする。
3. 対象 0 と対象 1 を入れ替えて、手順 1.、2. をおこなう。

5 つ目の知識 KNO-5 は、未来の視覚重畳関係と未来の運動特徴から、現在の運動特徴を振り返り推定するための知識である。“KNO-5: 物体 X が物体 Y を隠している場合、物体 X が十分動いたときにも片側遮蔽であれば、物体 Y は物体 X に随伴する運動をしてきた” とする。KNO-5 が成立するための前提条件は、対象が重要な見え隠れをする近くの時刻で認識可能であること、そして観察対象が消滅しないことである。

KNO-5 を適用するアルゴリズムは、次のようである。

1. 片側遮蔽の時区間  $\mathbb{T}_{K5A,i} \in \mathbb{T}_{All}$  を探索する。つまり

$$\begin{aligned} S_H(T_{init}(\mathbb{T}_{K5A,i})) &= \dots = S_H(T_{fin}(\mathbb{T}_{K5A,i})) \\ &= \text{“One-way (Obj-0)”} \end{aligned}$$

2. 各  $\mathbb{T}_{K5A,i}$  について、対象 0 が始めの場所  $\bar{x}^0(T_{init}(\mathbb{T}_{K5A,i}))$  から十分動いたと初めて分かる時刻  $T_{N0}$  を探索する。ただし、 $s_M^0(T_{N0}) = \text{“Yes”}$ 、 $\|\bar{x}^0(T_{N0}) - \bar{x}^0(T_{init}(\mathbb{T}_{K5A,i}))\| > B$  とし、 $B$  は正の定数とする。さらにその後、対象 0 が動きを停止する ( $s_M^0(T_{N1}) = \text{“No”}$ ) 時刻  $T_{N1}$  を探索する。
3. 2. の  $T_{init}(\mathbb{T}_{K5A,i})$  を  $T_{N1}$  に代えて同様に探索し、対象 0 が対象 1 を隠しながら動いている時区間  $\mathbb{T}_{K5B,i} = \{[T_{init}(\mathbb{T}_{K5A,i}), N0], [N0, N1], \dots\}$  を決定する。
4. 3. まで求めた時区間  $\mathbb{T}_{K5B,i}$  について、運動随伴性と対象 1 の運動情報の書き換えをおこなう。つまり  $S_C(T_B) = \text{“Yes”}$ 、 $s_M^1(T_B) = \text{“Yes”}$  ( $T_B \in \mathbb{T}_{K5B,i}$ ) とする。
5. 対象 0 と対象 1 を入れ替えて、手順 1. から 4. までをおこなう。

## 4 時系列メタ情報からの作用・機能推定

前章までに計算されたメタ情報を元に、どの時刻にどの作用がなされているかを計算する。道具を下へ滑りこませる、あるいは道具で包んで持ち上げることが無いタスクだとすれば、二つの対象のメタ情報のみを用いて 1) 物体を物体に載せて移動する、2) 物体を物体に入れて移動する、という作用が推定できる。ただし、例外的作業についても、教示者がどちらの対象を掴んでいるかという情報を付加することで分類可能であると考えられる。

まず、各対象が動きはじめる時刻集合  $\mathbb{T}_{start}$  と、動きが終わる時刻集合  $\mathbb{T}_{stop}$  を探索する。

$$\mathbb{T}_{start}^k = \{T_a \mid s_M^k(T_{a-1}) = \text{“No”}, s_M^k(T_a) = \text{“Yes”}\} \quad (5)$$

$$\mathbb{T}_{stop}^k = \{T_b \mid s_M^k(T_{b-1}) = \text{“Yes”}, s_M^k(T_b) = \text{“No”}\} \quad (6)$$

次に、何かに載せられる、あるいは何かに入れられる作用がおこなわれている時刻を探索する。何かに載せられる時刻  $\mathbb{T}_{LOAD}^k$  は次のように、

$$\begin{aligned} \mathbb{T}_{LOAD}^k &= \{T_c \mid T_c \in \mathbb{T}_{stop}^k, S_D(T_c) = \text{“Near”}, \\ S_H(T_c) &= \text{“One-way (Obj-k)”}, S_C(T_{c-1}) \neq \text{“Yes”}\} \quad (7) \end{aligned}$$

何かに入れられる時刻集合  $\mathbb{T}_{BOX}^k$  は次のように計算できる。

$$\begin{aligned} \mathbb{T}_{BOX}^k &= \{T_d \mid T_d \in \mathbb{T}_{stop}^k, S_D(T_d) = \text{“Near”}, \\ S_H(T_d) &= \text{“One-way (Obj-}\bar{k}\text{)”}, S_C(T_{d-1}) \neq \text{“Yes”}\} \quad (8) \end{aligned}$$

ただし、対象  $\bar{k}$  は対象  $k$  でない対象を指すものとする。

続いて、対象が他の対象に隠れながら動いている時区間を推定する。動き始める時刻集合  $\mathbb{T}_{start*}$  と、動き終わる時刻集合  $\mathbb{T}_{stop*}$  は、以下のように計算できる。

$$\begin{aligned} \mathbb{T}_{start*}^k &= \{T_e \mid T_e \in \mathbb{T}_{start}^k, S_D(T_e) = \text{“Near”}, \\ S_H(T_e) &= \text{“One-way (Obj-k)”}, S_C(T_{e+1}) = \text{“Yes”}\} \quad (9) \\ \mathbb{T}_{stop*}^k &= \{T_f \mid T_f \in \mathbb{T}_{stop}^k, S_D(T_f) = \text{“Near”}, \\ S_H(T_f) &= \text{“One-way (Obj-k)”}, S_C(T_{f-1}) = \text{“Yes”}\} \end{aligned}$$

ただし、 $\mathbb{T}_{start*}$  あるいは  $\mathbb{T}_{stop*}$  に含まれる時刻では、他の対象に隠されている対象の場所が分からない。しかし、隠している方の対象では場所が分かることを利用し、以下のように場所の伝達をおこなう。

$$\mathbf{x}^k(T_n) = \mathbf{x}^{\bar{k}}(T_n), \text{ if } s_V^k(T_n) = \text{“Yes”}$$

$$(T_n \in \mathbb{T}_{start*}^k \text{ or } T_n \in \mathbb{T}_{stop*}^k, k = 0, 1) \quad (10)$$

さらに、各対象の動き始めの時刻、あるいは止まる時刻について、それがロボットが予め知っている場所であったならば、その場所を表す記号で置き換える（スナップする）。これは細かな場所の差を吸収し、シンボリックなタスクプランニングを可能にするものである。後述の実験では、既知の家具の上面から距離が一定以内の場合に、目的地がその家具であったとしてスナップする。

未知の面であった場合には、家具あるいは平面認識をおこない、その場所を覚えるといった作業が必要となる。ただし物体の移動が主な目的である場合、空中での一時停止は省略して良いと考えられるため、後述の実験でも略すものとした。さらに、ある場所から同じ場所へ移動するという、繰り返しの動きも消去する。

最後に、各対象の道具としての機能を推定する。 $\mathbb{T}_{LOAD}^k$ 、 $\mathbb{T}_{BOX}^k$  の時刻が存在する場合、それぞれ対象  $\bar{k}$  には上にものを載せることのできる機能“MAT”、中にものを入れることのできる箱のような機能“STORAGE”が存在すると推定できる。さらにその後、動きが観測された場合には、 $\bar{k}$  には上にものを載せて運ぶことのできる機能“TRAY”、中にものを入れて運ぶことのできる機能“CONTAINER”があ

ると推定できる。

このように計算される道具的機能とタスク目的は、シンボリックなタスクプランニングへ利用可能である [9]。

## 5 実験

3章で述べた遮蔽下での対象の状態推定法、そして4章で述べた片付けタスクにおける道具の作用と機能の推定法を評価するため、意図的に抜けがある物体追跡結果を作成し、V-FESの入力とする実験をおこなった。

実験に使用した観測データセットは、Fig.4に示したOBS-1とOBS-2の二セットである。OBS-1では、コップと皿を観察対象とする。コップは皿に重ねられ、両観察対象ともにキッチンへ運ばれる。OBS-2ではボールと箱を観察対象とする。ボールが箱に入れられ、箱が棚へと収納されるが、箱を運ぶ際に両観察対象ともに棚に隠れ、全く画像内に写らない時区間が存在する (Fig.4, OBS-2の12.0[sec])。

### 5.1 抜けのある追跡データ

一般に画像中の対象探索では、観測可能な対象の特徴が多いほど探索結果は良好となる。対象の特徴が均一であれば、対象の観測可能な領域のサイズに依存する。

そこで、時刻  $t$  に対象  $k$  が認識される確率  $P^k(t)$  を、対象  $k$  の領域が隠れない場合に画像中で取り得る面積  $A_{all}^k(t)$  と、実際に観測可能な領域の面積  $A_{rec}^k(t)$  を用いて、次式のように定めた。

$$P^k(t) = \begin{cases} 0, & \text{if } A_{rec}^k \leq A_{min} \text{ or } \frac{A_{rec}^k(t)}{A_{all}^k(t)} \leq r_{min} \\ \frac{P_{max}}{1-r_{min}} \left( \frac{A_{rec}^k(t)}{A_{all}^k(t)} - r_{min} \right), & \text{otherwise} \end{cases} \quad (11)$$

ただし、 $A_{min}$  は認識に最低限必要となる観測可能領域のサイズ、 $P_{max}$  は対象領域全てが観測可能だった際の認識成功率、 $r_{min}$  は認識に最低限必要となる対象の観測可能領域の割合を表す定数で  $0 < r_{min} < 1$  とする。

$P_{max} = 0.7$ ,  $r_{min} = 0.5$ ,  $A_{min} = 400[\text{pixel}]$  として、対象の探索に成功するフレームをシミュレートした結果を、Fig.5に示す。OBS-1ではコップを重ねたときの皿、OBS-2では箱に入れられたときのボールと、棚に隠れたときの箱が認識されていないことが分かる。以降の実験は、各観察対象について、探索成功フレームについてのみ、正しい世界座標系における三次元位置  $x^k(t)$  が認識されたものとしておこなう。

### 5.2 時系列メタ情報の推定

3章で述べた手法を用い、隠れたメタ情報の推定をおこなった。まず3.2までで計算されたメタ情報が、Fig.6である。この段階では視覚重畳関係と距離はほぼ求まっておらず、動きの有無も半分程度しか推定できていない。これに対し、KNO-1からKNO-3を用いて修正し、運動随伴性を追加した後のメタ情報列がFig.7である。KNO-1を用いることで、コップと皿、ボールと箱の視覚重畳関係が推定されている。またKNO-2によって、棚に隠れている最中の箱とボールの視覚重畳関係が推定されている。さらにKNO-3を用いることで、OBS-1、OBS-2ともに近接関係が推定されている。続いてKNO-1からKNO-3までを繰り返し利用することで、OBS-2で最後まで、近接関係と視覚重畳関係が推定できている。最後に、KNO-4とKNO-5を用いて修正した情報列がFig.8である。KNO-4によって見えない箱の動きが推定され、KNO-5を用いることで、隠された状態での皿とボールの動き、そして運動随伴性が推定された。

### 5.3 片付けタスクにおける道具の作用・機能推定

続いて、4章で述べたタスク目的・機能推定法を適用した結果をFig.9に示す。ただし、OBS-1では机とキッチン、

OBS-2では机と棚に対象がスナップ可能であるようにした。

結果、OBS-1では皿にコップを重ねた状態でキッチンに置くこと、OBS-2ではボールを箱に入れて棚に置くことがタスクの目的であると推定された。さらに、皿には上に物を重ねて良いことと重ねた状態で運んで良いこと、箱には中に物を入れて良いこととその状態で運んで良いことが正しく推定された。

## 6 結言

本稿では、ロボットが道具を用いたタスク観察学習をおこなう際、観察対象同士に遮蔽が発生した場合、あるいは観察対象を見失った場合にも、各時刻で発生している作用と、片付けタスクの学習・計画へ必要となるタスク目的および道具的機能を推定する手法“V-FES”について述べた。提案手法では、まず観察対象同士の運動特徴と視覚重畳関係というメタ情報の時間的推移を推定する。情報同士を結びつける五つの知識を利用して、遮蔽下にある対象のメタ情報も推定し、メタ情報の組み合わせによって作用を推定する。特に、物体を物体で運ぶ際の運動随伴性の推定には、未来のメタ情報を利用する振り返り推定(3.5節)が重要であった。さらに、対象の認識結果を通知しあう機構(4章)が、遮蔽された対象の移動の目的地を推定するために有用であった。本稿の実験では、欠けのある仮想追跡結果を入力とし、片付けタスクでの本手法の有用性を評価した。実際の視覚処理手法と組み合わせ、実環境での片付けタスク学習をおこなった結果については文献[9]を参照されたい。

## 文献

- [1] Y. Kuniyoshi, M. Inaba and H. Inoue, “Learning by Watching: Extracting Reusable Task Knowledge from Visual Observation of Human Performance,” *IEEE Transactions on Robotics and Automation*, Vol. 10, No. 6, pp. 799–822, 1994.
- [2] K. Ikeuchi and T. Suehiro, “Toward an Assembly Plan from Observation: Task Recognition With Polyhedral Objects,” *IEEE Transactions on Robotics and Automation*, Vol. 3, Issue. 3, pp. 368–385, 1993.
- [3] K. Ito and S. Sakane, “Robust View-based Visual Tracking with Detection of Occlusions,” *Proc. of the 2001 IEEE International Conference on Robotics and Automation*, pp. 1207–1213, 2001.
- [4] K. Nagahama, K. Yamazaki, K. Okada and M. Inaba, “Hierarchical Estimation of Multiple Objects from Proximity Relationships Arising from Tool Manipulation,” *Proc. of the 2012 IEEE-RAS International Conference on Humanoid Robots*, pp. 666–673, 2012.
- [5] 今井, 金子, “対象の自律的切替機構を持つパーティクルフィルタによる遮蔽環境下での視覚追跡,” 電子情報通信学会技術研究報告. NC, ニューロコンピューティング, Vol. 106, No. 500, pp. 31–36, 2007.
- [6] H. Grabner, J. Matas, L. Van Gool and P. Cattin, “Tracking the Invisible: Learning Where the Object Might be,” *Proc. of the 2010 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1285–1292, 2010.
- [7] D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” *Proc. of the 1999 IEEE International Conference on Computer Vision*, pp. 1150–1157, 1999.
- [8] S. Hinterstoisser, S. Holzer, C. Cagniard, S. Ilic, K. Konolige, N. Navab and V. Lepetit, “Multimodal Templates for Real-Time Detection of Texture-less Objects in Heavily Cluttered Scenes,” *Proc. of the 2011 IEEE International Conference on Computer Vision*, pp. 858–865, 2011.
- [9] 長濱, 宗玄, 土永, 山本, 矢口, 稲葉, “生活支援ロボット HSR のための遮蔽情報推定に基づく片付けタスク視覚学習法,” 第14回 SICe システムインテグレーション部門講演会講演概要集, 1K1-2, 2013, to appear.
- [10] R. E. Fikes and N. J. Nilsson, “STRIPS: A NEW APPROACH TO THE APPLICATION OF THEOREM PROVING TO PROBLEM SOLVING,” *Proc. of the 1971 International Joint Conference on Artificial Intelligence*, pp. 608–620, 1971.
- [11] K. Nagahama, K. Yamazaki, K. Okada and M. Inaba, “Manipulation of Multiple Objects in Close Proximity Based on Visual Hierarchical Relationships,” *Proc. of the 2013 IEEE International Conference on Robotics and Automation*, pp. 1295–1302, 2013.

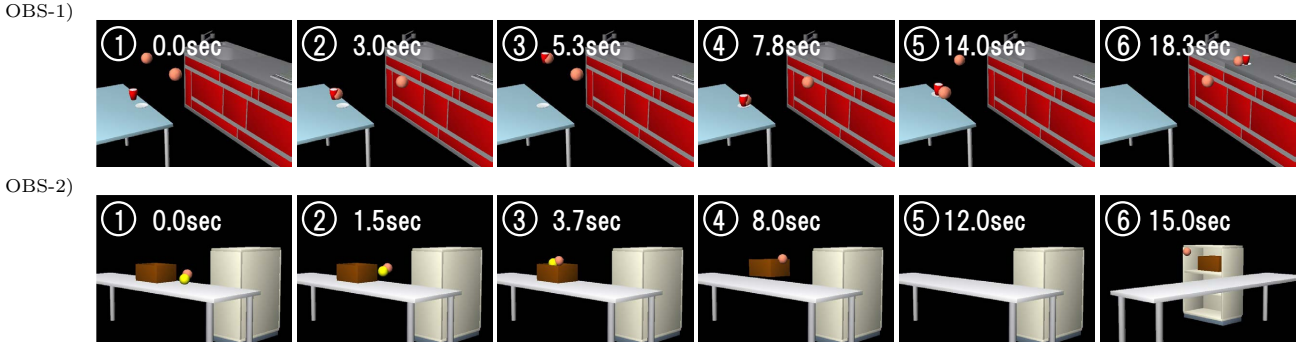


Fig. 4: Observed tasks: OBS-1 and OBS-2

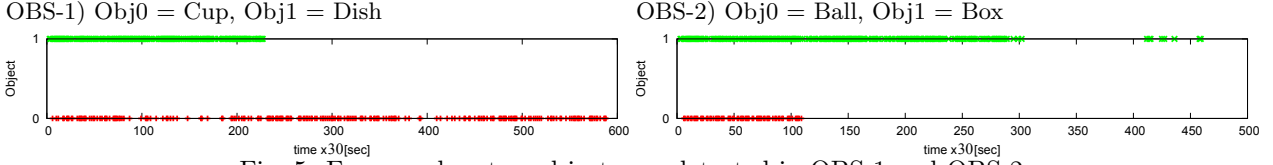


Fig. 5: Frames when two objects are detected in OBS-1 and OBS-2

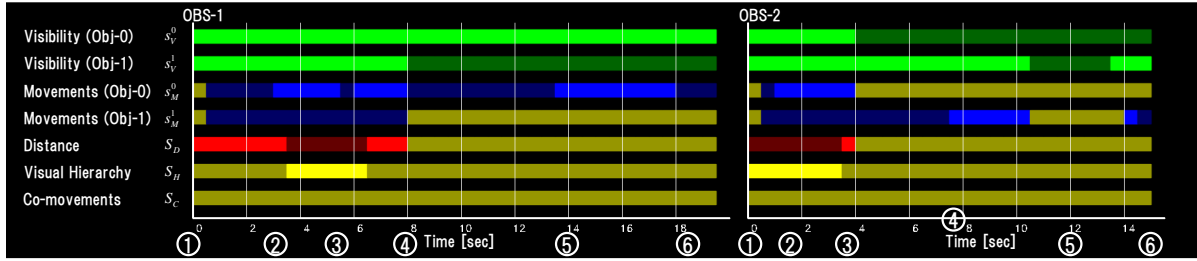


Fig. 6: Initial meta-information graphs. The circled numbers correspond the numbers in Fig.4

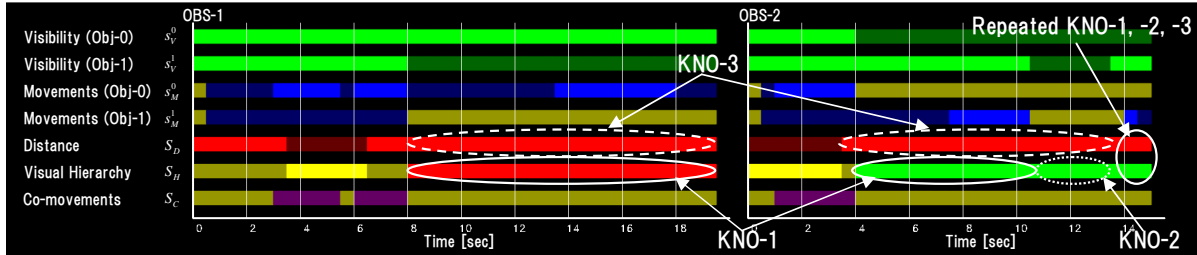


Fig. 7: Meta-information graphs after using KNO-1, 2, 3 repeatedly and the initial estimation of co-movements

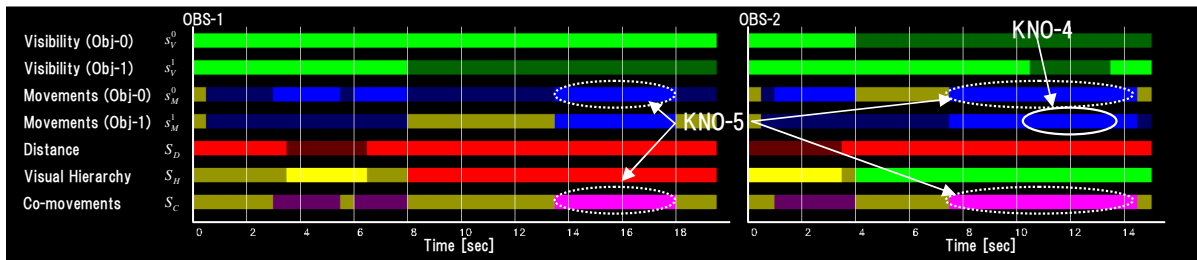


Fig. 8: Final meta-information graphs after using KNO-4 and KNO-5

OBS-1)

```

1 * OBJECT0
2 AT TIME=2.5[sec], MOVED_FROM PLACE= TABLE,
3 AT TIME=8[sec], LOADED_ONTO OBJECT1,
4 (AT TIME=13[sec], START being TRANSPORTED, by OBJECT1)
5 (AT TIME=18[sec], STOP being TRANSPORTED, by OBJECT1)
6 * OBJECT1
7 AT TIME=13[sec], MOVED_FROM PLACE= TABLE,
8 AT TIME=18[sec], MOVED_TO PLACE= KITCHEN,
9 ;; TOOL INFORMATION:
10 * OBJECT1 CAN BE USED AS MAT
11 * OBJECT1 CAN BE USED AS TRAY
12 ;; TASK GOAL:
13 ON OBJECT0 OBJECT1
14 AT OBJECT1 KITCHEN

```

OBS-2)

```

1 * OBJECT0
2 AT TIME=0.5[sec], MOVED_FROM PLACE= TABLE,
3 AT TIME=4[sec], PUT_INTO OBJECT1,
4 (AT TIME=7[sec], START being TRANSPORTED, by OBJECT1)
5 (AT TIME=14.5[sec], STOP being TRANSPORTED, by OBJECT1)
6 * OBJECT1
7 AT TIME=7[sec], MOVED_FROM PLACE= TABLE,
8 AT TIME=14.5[sec], MOVED_TO PLACE= SHELF,
9 ;; TOOL INFORMATION
10 * OBJECT1 CAN BE USED AS STORAGE
11 * OBJECT1 CAN BE USED AS CONTAINER
12 ;; TASK GOAL:
13 IN OBJECT0 OBJECT1
14 AT OBJECT1 SHELF

```

Fig. 9: List of estimated functions using OBS-1 and OBS-2