LI-007

身体の軌跡情報の解析に基づいた動作と動作対象の統合的認識

Integrated Recognition of Human Actions and Objects Based on Analysis of Body Movements

齊藤 雅紘 † 小島 篤博 ‡ 北橋 忠宏 § 福永 邦雄 †
Masahiro Saitou Atsuhiro Kojima Tadahiro Kitahashi Kunio Fukunaga

1. まえがき

近年，ロボットビジョンの実現を目指し，シーン中の人間の行動や物体の認識といった映像理解に関する研究が盛んに行われているが，これら従来研究の多くは動作と物体を個別に認識するものであった．しかしながら，人間の動作の多くには動作の対象となる事物が存在し，こういった動作と動作対象の関連性に着目したいくつかの研究が報告されている [1][2]．本研究ではこれに加え，人間の動作の特徴がよく表れる身体の動きを解析することにより，動作と動作対象を統合的に認識する手法を提案する．認識の過程において，時間的な事象の変化を扱うことが可能である DBNs(Dynamic Bayesian Networks)[3] と，事象の依存関係などの知識を用いることが可能であるフレーム表現とを相補的に組み合わせることで，知識と時系列を考慮に入れた認識が可能になる．

2. 提案手法の概要

2.1 人物動作と物体の階層モデル

一般に対象物を伴う動作には，動作の抽象度に対応した対象物の範疇を考えることができる．例えば，飲食を包含する抽象的な「摂食動作」には,「飲食物」という抽象的な物体を対応付けることが可能である．本研究では，フレーム表現を用いて手の動作と物体とを階層的にモデル化し (図 1)，抽象度に応じた関連付けを行う．手の動作の階層モデルの各ノードは，スロットに IS-A(上位-下位の関係)，OBJECT(動作対象)，PRECOND(事前条件)，EFFECT(事後条件) などを持つ．このうち OBJECT が物体の階層モデルのノードに対応している．物体の階層モデルのノードでは，他に FUNCTION(機能) などがある．

2.2 処理の流れ

本手法では **2.1** で定義した階層モデルを用い，身体の動きを解析することによって動作と動作対象を統合的に認識する．まず，入力データより抽出した特徴量を用い，**3.** で述べる確率的手法によって動作を「机上動作 : s^1」「摂食動作 : s^2」「垂直板上動作 : s^3」「その他の動作 : s^0」の 4 種類に分類する．ここで，確率の高い動作をその時刻における動作の候補とする．しかしながら，これより以下の動作は，対象物 (OBJECT) が判別できないと詳細化できないため，フレーム内のスロットに記述されている対象物に関する条件を探索する．そして，物体の階層構造における該当フレームを参照し，これを下方へたどって対象物を具体化する．これにより，対象物の種類が判別可能であれば，並行して動作の階層を詳細化することができる．以上のように，2 つの階層を交互にたどることにより,「食べ物」を「食べる」といったような具体的な動作と動作対象を統合的に認識する．

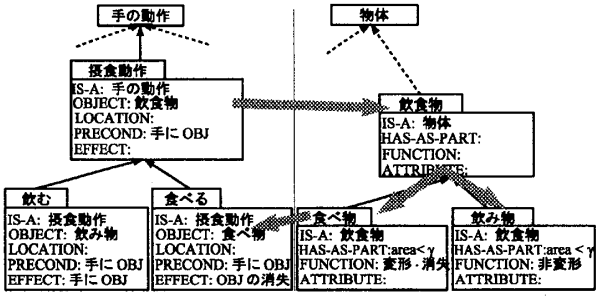

図 1: 動作と物体の階層構造の一部

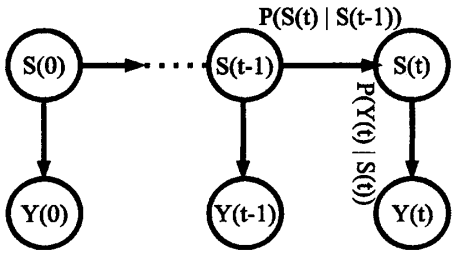

図 2: DBNs の基本モデル

3. 確率的手法を用いた動作分類

3.1 動作特徴の抽出

本研究では動作の分類に，時系列を扱える DBNs を用いる (図 2)．映像，距離データ，マーカーによる手と頭部の 3 次元位置を入力し，これらより DBNs の観測値 Y として右手と左手の位置 (R_p,L_p)，手と頭部との距離 (D_{rh},D_{lh})，手の移動方向 (R_d,L_d) をそれぞれ 3~4 種類に量子化した動作特徴を抽出する．また，机や黒板などの物体の手の動作への関与を考慮するために，手の付近での水平面の有無 (VSO_r,VSO_l) および垂直面の有無 (HSO_r,HSO_l) も特徴量として抽出する．

3.2 DBNs を用いた動作分類

時刻 t において得られた特徴量を DBNs の観測値 $Y_{(t)}$ とし，各時点におけるそれぞれの動作の事後確率を計算する．時刻 t までに得られた観測値の系列を $Y_{1:t}$ とすると，式 (1) により事後確率 $P(S_{(t)}|Y_{1:t})$ を計算でき，確率が閾値以上であればその時点の動作の候補とする．

†大阪府立大学大学院 工学研究科, Graduate School of Engineering, Osaka Prefecture University
‡大阪府立大学 総合教育研究機構, Library and Science Information Center, Osaka Prefecture University
§関西学院大学 理工学部 情報科学科, School of Science and Technology, Kwansai-Gakuin University

$$P(S_{(t)}|Y_{1:t}) = \alpha P(Y_{(t)}|S_{(t)}) \sum_{S_{(t-1)}} P(S_{(t)}|S_{(t-1)}) P(S_{(t-1)}|Y_{1:t-1}) \quad (1)$$

ただし，$S_{(t)}$ は時刻 t において推定する隠れノード，$P(S_{(t)}|S_{(t-1)})$ は時刻 $t-1$ から時刻 t へ状態が遷移する確率，$P(Y_{(t)}|S_{(t)})$ は状態が $S_{(t)}$ である時に観測値が $Y_{(t)}$ となる確率である．$P(Y_{(t)}|S_{(t)})$ については，あらかじめ撮影してある学習用映像を用い，最尤法によって確率値を学習させておく．また α は正規化定数であり，$\sum_i s^i = 1$ とする．本研究では，右手と左手の動作を区別して認識するために 2 種類の DBNs を用いる．

4. 動作と動作対象の統合的認識

3. で得られた動作の分類結果を基に動作の対象物について検証し，対象物との関連性から下位概念のより詳細な動作を認識する．そのために，まず人物の肌領域と前景領域を確率的手法を用いて検出する．そして得られた前景領域のうち人物領域以外の孤立領域を物体領域として抽出する．

以下，分類結果が「摂食動作」である場合を例に説明する．認識の過程は図 1 に太線で示す．まず，確率的手法のみでは「摂食動作」の下位概念である動作の判定は困難であるので，スロット OBJECT に「飲食物」が関連づけられていることに注目し，物体の階層構造における該当フレームを参照する．そしてその下位概念である「食べ物」「飲み物」の各スロットに記述されている条件を，入力映像より得られる特徴を用いて検証する．本研究ではスロット FUNCTION にあるように「食べ物」である条件を「変形，消失するもの」とし，また「飲み物」については，缶などの容器であることを仮定して，条件を「変形しないもの」とする．

各スロットにおける条件を満たし，物体を「食べ物」と決定できたとする．ここで先述の「摂食動作」の下位概念として，以下に示す「食べる」のスロット OBJECT に「食べ物」が関連づけられているので，これより動作を「食べる」と決定できる．

食べる	
IS-A:	**摂食動作**
OBJECT:	**食べ物**
LOCATION:	
PRECOND:	**手に** *OBJ*
EFFECT:	*OBJ* **の消失**

以上のように各モデルの下位概念を双方向的に検証することにより，動作と動作対象を統合的に認識する．

5. 実験

提案手法の有効性を確認するため，登場人物が食べ物を机の上に置きその後食べるといったシーンを対象に実験を行った．実験の内容は，まず DBNs を用いて動作の分類を行い，次に得られた動作の分類結果を用いて動作対象に関する条件を検証することにより，下位概念の動作を認識するといったものである．なお，「その他の動作」以外の観測確率 $P(Y_{(t)}|S_{(t)})$ は 3.2 で述べた方法で学習させた．右手の動作の分類結果を図 3 に示す．縦軸は確率，横軸はフレーム数 (#) を表し，ここでは「机上動作」

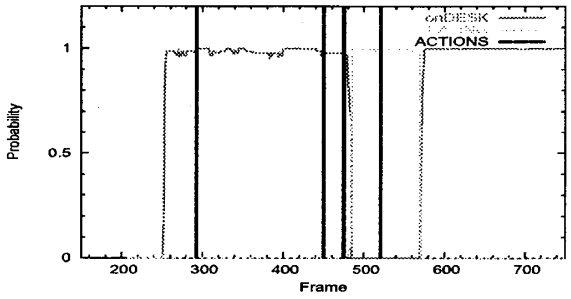

図 3: 右手の動作の分類結果

#338 PUT DOWN. (left hand)　　#521 EAT. (right hand)

図 4: 実験映像と認識結果の例

「摂食動作」の確率を表している．例えば，#250~#480 付近では机上動作，#480~#580 付近では摂食動作がそれぞれ生起していると判定できる．また，各動作が認識されたフレームを図 3 中に太い縦線で示してある．#521 では「摂食動作」の確率が高く，そこで手に持っている物体が消失したので動作対象を「食べ物」，動作を「食べる (右手)」と認識した (図 4)．同様に，#293 で「置いた (右手)」，#338 で「置いた (左手)」，#475 で「取った (右手)」の認識に成功している．一方，物体領域の誤抽出により動作を正しく認識できない箇所があった (#450)．

6. むすび

本稿では，手と頭部の動きを解析することにより人間の動作を分類し，その結果から動作と物体の階層モデルの下位概念を検証することで，より詳細な動作と動作対象を統合的に認識する手法を提案した．また，人物が飲む，食べるといった動作を行うシーンに対して実験を行い，提案手法の有効性を確認した．

参考文献

[1] 樋口未来，小島篤博，福永邦雄，“人間の動作と動作対象の関連性に基くシーンの統合的認識，” MIRU2004, pp.II-329-II-334, July2004.

[2] D. Moore, I. Essa, and M. Hayes, “Exploiting Human Actions and Object Context for Recognition Tasks, ” In *Proc. of IEEE International Conference on Computer Vision*, Corfu, Greece, March 1999.

[3] Kevin P. Murphy, “Dynamic Bayesian Networks: Representation, Inference and Learning, ” PhD thesis, University of California, 2002.