2I5-OS-08b-5

物体指向動作の心象と表象の確率的カテゴリゼーション

Probabilistic Categorization of Images and Symbols of Object-oriented Actions and Activities

渥美雅保
Masayasu Atsumi

創大・工・情報システム工学
Dept. of Information Systems Sci., Faculty of Eng., Soka University

This paper proposes a probabilistic categorization method of object-oriented actions and activities in which their visual appearance classes of motion and their semantic tags in a form of deep case triplets are organized into a probabilistic semantic network through the incremental probabilistic latent component analysis and actions are correlated with their contextual activities. Through experiments using video clips captured by a Kinect sensor, it is shown that the method can learn, recognize and infer object-oriented actions in context of activities.

1. はじめに

日常生活空間における人とロボットのインタラクションにおいて，人が何をしているのかをロボットが理解することはロボットが人を自律的に支援するうえで必要不可欠な機能である．人の動作には物体を用いるための物体への働きかけ動作が多くみられる．このとき，動作を理解するためには，動きとその動きが働きかける対象である物体を認識することが必要であり，動作と物体はペアで実世界における意味を形成する．本研究では，このような観点から捉えられる動作を「物体指向動作」と呼ぶ．これは，物体が可能な動作を規定して動作に意味を与えていると捉えれば，心理学におけるアフォーダンス，並びに情報工学におけるオブジェクト指向の世界観である．ところで，1 つ 1 つの動作はあるコンテキストの中で行われることが多い．例えば，フォークやスプーンを使う動作はそれら一連の動作からなる食事をするというコンテキストの中で多く行われる．本研究では，前者の 1 つ 1 つの動作を「アクション」，後者の一連の動作を「アクティビティ」と呼び，アクティビティがアクションのコンテキストを与えてアクションの認識を促進すると仮定する．これは，物体の認識がそのコンテキストが与えられることにより促進されるという知見 [Bar 04] と共通のものであり，物体認識におけるコンテキストの利用 [Atsumi 13] の拡張である．物体指向のアクション及びアクティビティは視覚的な動き特徴からクラス分けすることができ，それらの意味はクラスのラベル付けにより与えることができるが，一般に，それら視覚的特徴とその言語ラベルは一対一には関連付けられない．そこで本研究では，それらクラスを＜対象意味素 (target synset), 格 (case), 動作意味素 (motion synset) ＞の格 3 つ組でラベル付けし，視覚的動き特徴を表すクラスと言語的意味を与える格 3 つ組の意味素をノードとする確率的な意味ネットワークでアクション及びアクティビティの心象と表象を表現する．また，アクションとそのコンテキストを与えるアクティビティの関連をそれらの共起関係により表現する．

本論文では，物体指向動作の観測とそれへの格 3 つ組によるラベル付けが与えられて，アクション，及びアクティビティの視覚的動き特徴の心象を表すクラスとその格表現に基づく表象を確率的に関連付けた意味ネットワークを学習する手法，及びそれを用いたアクション及びアクティビティの認識と推論の手法を提案する．本手法の特徴は，次の 3 つである．第 1 に，アクション，及びアクティビティの視覚的な動き特徴を表すクラス集合をインクリメンタル確率潜在コンポーネント解析 (Incremental Probabilistic Latent Component Analysis, I-PLCA) により学習する点，第 2 に，アクション，及びアクティビティの視覚的な動き特徴を表すクラスとそれらの言語的意味を与える格 3 つ組の意味素の間の関連を確率的な意味ネットワークに獲得して，視覚レベルの動作認識と言語レベルの動作推論を融合している点，第 3 に，アクションとアクティビティの共起関係を求めて，それを用いてアクティビティをコンテキストとしたアクションの認識を実現している点である．

関連する研究として，木谷ら [Kitani 08] は，プリミティブ動作のカテゴリゼーションに，動作と関連する物体や背景の見えをコンテキストとして利用する方法を提案している．また，Yao ら [Yao 12] は，静止画像を対象として，物体と人の姿勢を相互にそれぞれの認識のコンテキストとして利用するモデルを提案し，さらに [Yao 13] では，同じく静止画像を対象として，物体に働きかける様々な姿勢を物体が有する様々な機能に発見的に対応づけるモデルを提案している．本研究では，動作だけでなく動作系列のカテゴリゼーションも扱い，物体に加えて動作系列を動作のカテゴリゼーションのコンテキストに利用している点が大きな違いである．また，物体に対する様々な動作の意味を格 3 つ組を用いて確率的に与えて推論に利用している点もこれら既存研究にない特徴である．

2. 物体指向動作の確率的カテゴリゼーション

2.1 提案手法の概要

人の動作を身体スケルトンのジョイント点の 3 次元座標の時系列としてキャプチャする．本研究では，両手による物体指向動作を扱うため，肩中心に対する両手の相対 3 次元座標の時系列を利用する．これら相対 3 次元座標は，Kinect センサーの RGB-D 画像から計算されるスケルトンのジョイント座標を用いて求めることが可能である．

両手の相対 3 次元座標の時系列から，両手の動き特徴量を次の手順により求める．まず，両手の相対 3 次元座標をある間隔で量子化し，量子化された相対位置とその変位の時系列を計算する．次に，それら時系列に対して，アクション，及びその系列であるアクティビティのアノテーションを，それらの開始フレームと終了フレーム，及び格 3 つ組を指定することにより付与する．そして，各アクション，及びアクティビティの相

連絡先: 渥美雅保，masayasu.atsumi@gmail.com

対位置と変位の時系列に対して，その時系列が表す動き特徴量を，肩中心を原点として身体周りの 3 次元空間をある大きさで分割したブロックごとの変位のヒストグラムとして求める．

アクションの確率的意味ネットワークの学習では，アクションの格 3 つ組付きヒストグラムの集合を入力として，動き特徴を表すクラス集合とそれらと格 3 つ組の意味素との確率ネットワークを求める．まず，アクションのヒストグラムの集合から，I-PLCA によりアクションクラス集合を求める．次に，アクションクラスと意味素の結合確率の計算に基づいて，アクションクラスと意味素のネットワークを生成する．アクティビティの確率的意味ネットワークの学習でも，同様に，アクティビティの格 3 つ組付きヒストグラムの集合を入力として，動き特徴を表すクラス集合とそれらと格 3 つ組の意味素との確率ネットワークが求められる．また，アクションとアクティビティの共起関係をアクションとアクティビティの格 3 つ組の確率から自己相互情報量を計算することにより求める．このアクション・アクティビティの共起関係づけられた確率的意味ネットワークを本論では ACTNET と呼ぶ．図 1 に ACTNET の構成を示す．

アクション及びアクティビティの認識と推論では，アクションのヒストグラムの系列入力に対して，それらアクションの格 3 つ組，及びアクティビティの格 3 つ組を求める．まず，各アクションのヒストグラムに対して，アクションクラスを確信度付きで求める．同時に，アクション系列のヒストグラムの和に対して，アクティビティクラスを確信度付きで求める．次に，アクションとアクティビティの格 3 つ組を，それら求められたクラスに基づく確率意味ネットワーク上での確率推論とアクションとアクティビティの共起関係を用いて求める．

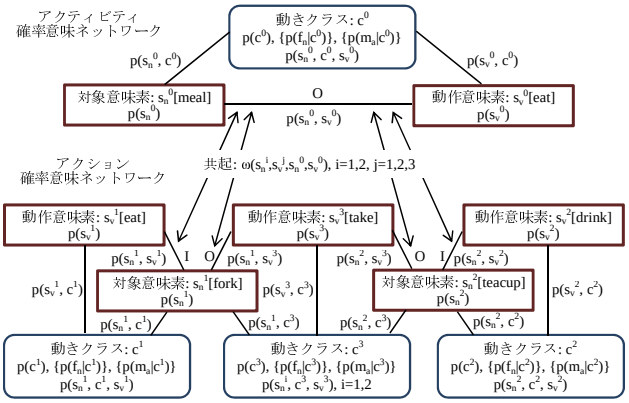

図 1: ACTNET の構成 (図中の記号は本文を参照のこと)

2.2 物体指向動作の動き特徴量

両手の量子化された 3 次元相対位置を $p^l = (p^l_x, p^l_y, p^l_z)$ と $p^r = (p^r_x, p^r_y, p^r_z)$，その変位を $d^l = (d^l_x, d^l_y, d^l_z)$ と $d^r = (d^r_x, d^r_y, d^r_z)$ とする．ここで，l は左手，r は右手を表し，変位は 2 フレーム間の量子化された相対位置の差で与えられる．また，相対位置と変位の時系列に付与される格 3 つ組を $\langle s_n[w_n], r, s_v[w_v]\rangle$ とする．ここで，w_n は動作の対象を表す名詞，s_n はその意味素，w_v は動作を表す動詞，s_v はその意味素で，意味素は日本語 WordNet[Isahara 08] の同義集合 (synset) により与えられる．また，r は格表記で，現状，対象格 (O)，道具格 (I)，場所格 (L[at|inside|around|above|below|beyond|from|to]) を与えることができる．アクションの時系列にはその格 3 つ組に加えて，それを含むアクティビティの格 3 つ組が付与される．このとき，格 3 つ組が付与されたアクション，またはアクティビティの相対位置と変位の時系列 $m = \{(p^l, d^l, p^r, d^r)_t\}$ に対して，その時系列が表す動き特徴のヒストグラムを次のように構成する．いま，肩中心を原点として身体周りの 3 次元空間を分割したブロックの集合を B，ブロックの数を $|B|$ とする．ブロック $b \in B$ に相対位置が含まれる時系列 m の部分系列を $m(b)$ とするとき，$m(b)$ の動き特徴を $m(b)$ に含まれる各要素の左手と右手それぞれの x，y，z 方向の変位が正か 0 か負かに応じてそれらを 27 個のビンに振り分けたヒストグラムとして構成する．そして，m の動き特徴のヒストグラム $h(m)$ を，これら $|B|$ 個の部分ヒストグラムを連結した $27 \times |B|$ 次元のヒストグラムで表現する．このヒストグラムは，身体周りでの両手の動きをとらえている．

2.3 物体指向動作の学習

ACTNET の学習は，アクションとアクティビティの確率的意味ネットワークの生成，並びにアクションとアクティビティの共起関係の設定によりなされる．アクションとアクティビティの確率的意味ネットワークは，それぞれアクションの格 3 つ組付き動きヒストグラム集合とアクティビティの格 3 つ組付き動きヒストグラム集合から生成される．いま，格 3 つ組 a が付いた動き時系列 m のヒストグラムを $h(m_a) = [h_{m_a}(f_1), \ldots, h_{m_a}(f_{|F|})]$，その集合を $H = \{h(m_a)\}$ とする．ここで，$f_i \in F$ はヒストグラムのビンで，ヒストグラムのサイズは $|F| = 27 \times |B|$ である．確率的意味ネットワークの生成は，格 3 つ組付き動きヒストグラム集合の I-PLCA を用いた確率的クラスタリング [Atsumi 12] による動きクラスの生成と，動きクラスと格 3 つ組の意味素との結合確率の計算に基づくネットワーク生成により遂行される．

I-PLCA による動きクラスの生成は，動きヒストグラム集合 $H = \{h(m_a)\}$ に対して，対数尤度

$$L = \sum_{m_a} \sum_{f_n} h_{m_a}(f_n) \log p(m_a, f_n) \tag{1}$$

を最大にする $p(m_a, f_n) = \sum_c p(c)p(m_a|c)p(f_n|c)$，即ちクラス確率分布 $\{p(c)|c \in C\}$，インスタンス確率分布 $\{p(m_a|c)|m_a \in M \times A, c \in C\}$，クラス特徴確率分布 $\{p(f_n|c)|f_n \in F, c \in C\}$，及びクラスの数 $|C|$ を推定することによりなされる．ここで，C はクラス集合，M は動き系列集合，A は格 3 つ組集合で，m_a は格 3 つ組 a を付与された動き系列 m，即ち動きのインスタンスである．これらの確率分布とクラス数をクラス分割を伴う Tempered EM アルゴリズムにより推定する．そのため，1 つまたはある少数のクラスから初めて，一定回数を上限とする EM の繰返し毎に各クラス内のばらつき度

$$\delta_c = \sum_{m_a} \left(\left(\sum_{f_n} \left| p(f_n|c) - \frac{h_{m_a}(f_n)}{\sum_{f'_n} h_{m_a}(f'_n)} \right| \right) \times p(m_a|c) \right) \tag{2}$$

を求めて，ばらつき度が最大のクラスを 2 分割する．この過程をすべてのクラスのばらつき度が閾値以下またはクラス確率が閾値以下となるまで繰り返してクラス数を確定し，その後 Temperd EM により確率分布を決定する．

確率的意味ネットワークは，動きクラスに関連付けられたクラス確率分布 $\{p(c)|c \in C\}$ とインスタンス確率分布 $\{p(m_a|c)|m_a \in M \times A, c \in C\}$ を用いて生成される．ネットワークのノードは，各クラスに対応したクラスノードと格 3

つ組の意味素に対応した意味素ノードからなり，意味素ノードは，動きインスタンス m_a の格 3 つ組 a が $\langle s_n[w_n], r, s_v[w_v]\rangle$ であるとき，その対象意味素 s_n と動作意味素 s_v に対して生成される．また，それら意味素ノードには名詞 w_n と動詞 w_v がそれぞれ登録される．各クラスノード $c \in C$ には，クラス確率 $p(c)$，クラス特徴確率分布 $\{p(f_n|c)|f_n \in F\}$，インスタンス確率分布 $\{p(m_a|c)|m_a \in M \times A\}$，及び式 (3) で与えられるクラス c が対象意味素 s_n と動作意味素 s_v でラベル付けされる結合確率が保持される．

$$p(s_n, c, s_v) = p(c) \quad \times \sum_{a=\langle s_n[*],*,s_v[*]\rangle} p(m_a|c) \tag{3}$$

ここで，$*$ は任意の単語，または格を表す．クラス c のノードと対象意味素 s_n のノードを結ぶリンクには結合確率 $p(s_n, c)$ が，クラス c のノードと動作意味素 s_v のノードを結ぶリンクには結合確率 $p(s_v, c)$ が，対象意味素 s_n のノードと動作意味素 s_v のノードを結ぶリンクには結合確率 $p(s_n, s_v)$ が付与される．また，対象意味素 s_n のノードと動作意味素 s_v のノードには，それぞれ $p(s_n)$ と $p(s_v)$ が保持される．これら確率は式 (4) で求められる．

$$\begin{gathered} p(s_n, c) = \textstyle\sum_{s_v} p(s_n, c, s_v), p(s_v, c) = \sum_{s_n} p(s_n, c, s_v) \\ p(s_n, s_v) = \textstyle\sum_c p(s_n, c, s_v) \\ p(s_n) = \textstyle\sum_c p(s_n, c), p(s_v) = \sum_c p(s_v, c) \end{gathered} \tag{4}$$

アクションとアクティビティの間の共起関係は，格 3 つ組 $\langle s_n[w_n], r, s_v[w_n]\rangle$ を持つアクションが格 3 つ組 $\langle s_n^0[w_n^0], r^0, s_v^0[w_n^0]\rangle$ を持つアクティビティに含まれるときに，アクションの対象意味素 s_n と動作意味素 s_v のペアとアクティビティの対象意味素 s_n^0 と動作意味素 s_v^0 のペアの間に設定される．いま，アクションの対象意味素 s_n と動作意味素 s_v のペア結合確率を $p(s_n, s_v)$，アクティビティの対象意味素 s_n^0 と動作意味素 s_v^0 のペアの結合確率を $p(s_n^0, s_v^0)$ とする．このとき，これらペアの間の共起度は式 (5) により定められる．

$$\omega(s_n, s_v, s_n^0, s_v^0) = \log \frac{p(s_n, s_v, s_n^0, s_v^0)}{p(s_n, s_v)p(s_n^0, s_v^0)} \tag{5}$$

ここで，結合確率 $p(s_n, s_v, s_n^0, s_v^0)$ は，式 (6) によりアクションの動きインスタンスから求められる．

$$p(s_n, s_v, s_n^0, s_v^0) = \sum_c \Big(p(c) \quad \times \sum_{a=\langle s_n[*],*,s_v[*]\rangle @ \langle s_n^0[*],*,s_v^0[*]\rangle} p(m_a|c) \Big) \tag{6}$$

ここで，$a = \langle s_n[*], *, s_v[*]\rangle @ \langle s_n^0[*], *, s_v^0[*]\rangle$ は m がアクション格 3 つ組 $\langle s_n[*], *, s_v[*]\rangle$ を持ち，かつそれを含むアクティビティが格 3 つ組 $\langle s_n^0[*], *, s_v^0[*]\rangle$ を持つことを表す．

2.4 物体指向動作の認識と推論

アクションとアクティビティの認識と推論では，与えられるアクション系列に対して，各アクションの動きヒストグラムを順次 ACTNET に渡して，各々のアクションとそれまでの系列が表すアクティビティ，即ちアクションとアクティビティの対象意味素と動作意味素を求める．そのために，まず，動きクラスを認識し，次に，動きクラスに関連付けられた意味素を推論する．その際，部分系列から推論されたアクティビティはアクションのコンテキストとして共起関係を介してアクションの推論に作用する．

アクションの動きヒストグラムまたはアクション系列の動きヒストグラムに対するアクションまたはアクティビティの動きクラスの認識は次のようになされる．いま，$h(m) = [h_m(f_1), \ldots, h_m(f_{|F|})]$ を動き m のヒストグラムとし，その分布を $\hat{h}(m) = [\hat{h}_m(f_1), \ldots, \hat{h}_m(f_{|F|})]$ とする．このとき，動きクラスの認識は，クラスノードの動きクラス c のクラス特徴確率分布とこの動き m のヒストグラム分布の類似度を式 (7) により計算し，類似度が大きい動きクラスを求めることによりなされる．この類似度は動きクラスの確信度として用いられる．

$$\beta(c, m) = 1 - \frac{\sum_{f_n} |p(f_n|c) - \hat{h}_m(f_n)|}{2} \tag{7}$$

確信度付き動きクラスが求まるとそれから対象意味素・動作意味素を確信度付きで推論することができる．アクションまたはアクティビティの動きクラスを c，その確信度を β とする．このとき，ACTNET のノードとリンクに保持されている確率を用いて，そのクラスノードからリンクされる意味素ノードの対象意味素 s_n，動作意味素 s_v，及びそれらのペアが，それぞれ確信度 $p(s_n|c) \times \beta$，$p(s_v|c) \times \beta$，$p(s_n, s_v|c) \times \beta$ で推論される．また，追加情報として対象意味素または動作意味素が与えられたとき，動作意味素または対象意味素が同様にそれぞれ確信度 $p(s_v|c, s_n) \times \beta$ と $p(s_n|c, s_v) \times \beta$ で推論される．また，アクションとアクティビティの共起度を用いてアクションとアクティビティを同時に確信度付きで推論することも可能である．クラス認識で求められたアクションクラス c の確信度を β，アクティビティクラス c^0 の確信度を β^0 とする．このとき，それらクラスからリンクされるアクションの対象意味素と動作意味素のペア (s_n, s_v) とアクティビティの対象意味素と動作意味素のペア (s_n^0, s_v^0) の組に対する確信度 $\beta(s_n, s_v, s_n^0, s_v^0|c, c^0)$ は式 (8) で計算される．

$$\begin{aligned} \beta(s_n, s_v, s_n^0, s_v^0|c, c^0) = p(s_n, s_v|c) \times p(s_n^0, s_v^0|c^0) \\ \times(\beta + \beta^0)/2 + \lambda \times \omega(s_n, s_v, s_n^0, s_v^0) \end{aligned} \tag{8}$$

ここで，λ は共起係数である．また，アクティビティの意味素ペアが (s_n^*, s_v^*) と定められた場合，アクションの意味素ペア (s_n, s_v) が確信度 $\beta(s_n, s_v, s_n^*, s_v^*|c, c^0)$ で推論される．

3. 実験

3.1 実験枠組

物体指向動作の ACTNET への学習，及び ACTNET を用いた認識と推論の評価を，Kinect センサーを用いてキャプチャしたビデオクリップを用いて行った．身体スケルトンから得られる両手の 3 次元相対位置は $1cm$ 間隔で量子化され，また，それら相対位置はフレームレートが約 $30fps$ になるように補間された．本論文では，アクティビティのアノテーションとして格 3 つ組<食事, を, 食べる>と<イラスト, を, 描く>の 2 つが付けられた動き系列をビデオクリップから抽出して用いた実験の結果を示す．アクティビティ<食事, を, 食べる>には，3 つの物体とそれらに対する動作からなる計 9 個のアクション，アクティビティ<イラスト, を, 描く>には 2 つの物体とそれらに対する動作からなる 7 個のアクションが含まれる．アクションの総数は 16 個である．表 1 にアクティビティとアクションのアノテーションに用いた格 3 つ組を示す．また，図 2 に，格 3 つ組 < ティーカップ, で, 飲む > が付与されたアクションに含まれるいくつかの量子化された動きを示す．動きのヒストグラム化における身体周りのブロック分けは，身体の近傍の前方と側方をそれぞれ 1 辺 30cm の 9 ブロック，その外側の前方と側方をそれぞれ大きく 9 ブロックと 8 ブロッ

表 1: アクティビティとアクションのアノテーションに用いた格 3 つ組

アクティビティ	食事, を, 食べる:<07573696-n[meal],O,01166351-v[eat]>)	イラスト, を, 描く:(<03561345-n[illustration],O,01684663-v[paint]>
アクション	フォーク, を, 持つ:<03382948-n[fork],O,01216670-v[take]> フォーク, で, 食べる:<03382948-n[fork],I,01166351-v[eat]> フォーク, を, 置く:<03382948-n[fork],O,01494310-v[put]> ティーポット, を, 持つ:<04398044-n[teapot],O,01216670-v[take]> ティーポット, で, 注ぐ:<04398044-n[teapot],I,02070296-v[pour]> ティーポット, を, 置く:<04398044-n[teapot],O,01494310-v[put]> ティーカップ, を, 持つ:<04397452-n[teacup],O,01216670-v[take]> ティーカップ, で, 飲む:<04397452-n[teacup],I,01170052-v[drink]> ティーカップ, を, 置く:<04397452-n[teacup],O,01494310-v[put]>	ノート, を, 持ち出す:<06415419-n[notebook],O,02311387-v[take-out]> ノート, を, 開く:<06415419-n[notebook],O,01346003-v[open]> ノート, を, 閉じる:<06415419-n[notebook],O,01291941-v[close]> ノート, を, 戻す:<06415419-n[notebook],O,01308381-v[put-back]> ペンシル, を, 持つ:<03908204-n[pencil],O,01216670-v[take]> ペンシル, で, 描く:<03908204-n[pencil],I,01684663-v[paint]> ペンシル, を, 置く:<03908204-n[pencil],O,01494310-v[put]>

((-32,-38,33),(0,0,-1), (4,-15,30),(0,1,-1))　((-29,-37,32),(0,0,0), (9,-8,20),(0,0,0))　((-29,-37,32),(0,0,0), (11,-8,20),(-1,0,0))

図 2: アクションに含まれる両手の動きの符号化の例

ク，後方を 1 つのブロックとする．これよりブロック数は 36 となり，動きヒストグラムの次元は 972(= 27 × 36) 次元である．実験で用いたパラメータは，クラス分割におけるばらつき度の閾値とクラス確率の閾値がともに 0.1，アクションとアクティビティの共起係数が 0.2 である．

3.2 実験結果

物体指向動作が学習された ACTNET の構成を表 2 に示す．ここで，クラス数は I-PLCA のクラス分割により自動的に決められている．図 1 は，この ACTNET の構成の一部である．2 つのクラスノードからなるアクティビティネットワークの詳細な学習結果を図 3 に示す．

表 2: ACTNET の構成

	クラス数	対象意味素数	動作意味素数	意味素ペア数
アクティビティ	2	2	2	2
アクション	16	5	10	16

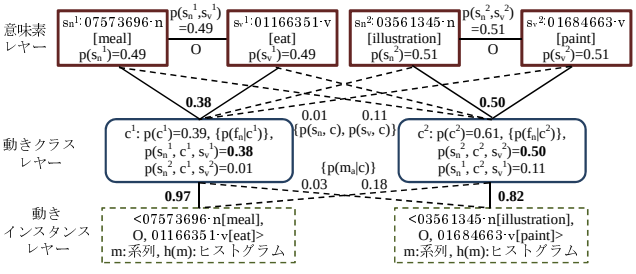

図 3: ACTNET のアクティビティサブネットワーク

表 3 に，正解付きのアクション系列に対する ACTNET による認識・推論の評価結果を示す．アクティビティの正解率は 100% であるが，アクティビティとの共起を利用せずに独立にアクションの認識・推論を行った場合の正解率は 75% であった．一方，アクティビティとの共起を利用した場合，アクションの正解率は 93.8% に上昇した．また，物体が何かの追加情報が得られた場合の正解率は 93.8% であった．アクションの次善解までの正解率は，共起の有無に関わらず 93.8%，さらに物体が何かの追加情報が得られた場合は 100% であった．

表 3: 認識・推論結果

アクティビティ正解率	100%
アクション正解率 (共起なし)	75.0%
物体がわかったときのアクション正解率 (共起なし)	93.8%
アクション正解率 (共起あり)	81.3%
物体がわかったときのアクション正解率 (共起あり)	93.8%

4. おわりに

本論では，物体指向のアクションとそのコンテキストを与えるアクティビティに関して，それらの視覚的動き特徴を表すクラスとその解釈を与える格 3 つ組の意味素とからなる確率的意味ネットワーク ACTNET を学習する手法と，それを用いてアクションとアクティビティを同時に関連付けて認識・推論する手法について述べた．そして，Kinect センサーでキャプチャした両手の物体指向動作のビデオクリップを用いて，学習された ACTNET によりアクションとアクティビティの認識が可能なこと，特に，アクション認識のあいまいさが追加情報を用いた推論により解消すること，コンテキストを与えるアクティビティとの共起によりアクションの認識が向上することを示して，本手法の有用性を明らかにした．

謝辞　本研究の一部は科学研究費補助金 (課題番号 23500188) の支援のもとに実施された．

参考文献

[Bar 04] Bar, M.: Visual Objects in Context, Nature Reviews Neuroscience, Vol.5, pp.617-629 (2004)

[Atsumi 13] Atsumi, M.: Object Categorization in Context based on Probabilistic Learning of Classification Tree with Boosted Features and Co-occurrence Structure, Lecture Note in Computer Science, Vol.8033, pp.416-426 (2013)

[Kitani 08] Kitani, K.M., Okabe, T. and Sato, Y.: Discovering Primitive Action Categories by Leveraging Relevant Visual Context, Proceedings of the IEEE International Workshop on Visual Surveillance (2008)

[Yao 12] Yao, B. and Fei-Fei, L.: Recognizing Human-object Interactions in Still Images by Modeling the Mutual Context of Objects and Human Poses, IEEE Trans. on Pattern Analysis and Machine Intelligence 34 (9) pp.1691-1703 (2012)

[Yao 13] Yao, B., Ma, J. and Fei-Fei, L.: Discovering Object Functionality, Int. Conf. on Computer Vision 2013 (2013)

[Isahara 08] Isahara, H. *et. al.*: Development of Japanese WordNet, 6th Int. Conf. an Language Resources and Evaluation, pp.2420-2423 (2008)

[Atsumi 12] Atsumi, M.: Learning Visual Categories based on Probabilistic Latent Component Models with Semi-supervised Labeling, GSTF Int. J. on Computing, Vol.2, No.1, pp.88-93 (2012)