情報処理学会論文誌　Vol. 51　No. 8　1413–1427 (Aug. 2010)

機能と視覚的情報の関係性に基づく物体の概念モデル

中村友昭[†1]　長井隆行[†2]

本論文では，視覚特徴と機能の関係に基づいた物体の概念モデルを提案する．物体は，使用目的や用途が存在しており，それらを満たすための機能を有している．そこで，道具が作用する対象物の道具使用前後の変化を機能と定義し，これをカメラで観測することで機能概念を学習する．さらに，この機能と視覚特徴の関係性により物体概念の学習を行う．視覚的特徴は，SIFT（Scale Invariant Feature Transform）を用いて表現する．物体概念はグラフィカルモデルによって表現されており，教師なしで物体概念を学習することができる．さらに，学習されたグラフィカルモデルを用いることで，未観測情報の予測を行うことが可能となる．たとえば，提案モデルでは，視覚特徴のみから機能を予測することが可能である．本論文では，実際に 49 個の道具を用いて実験を行い，提案モデルの有効性を示す．

Object Concept Model Based on the Relationship between Functions and Visual Information

Tomoaki Nakamura[†1] and Takayuki Nagai[†2]

This paper discusses a computational model for object concept formation. We propose a model of the object concept based on the relationship between appearance and functions. Implementation of the proposed framework using Bayesian Network is presented. At this point we need an explicit definition of the object function. In the proposed model, each function is defined as certain changes in a work object caused by the object. Therefore, each function is represented by a feature vector which quantifies the changes in the work object. Then the function is abstracted from these feature vectors using the Bayesian learning approach. The system can form object concepts by observing human tool use based on the abstract function and shape information. Furthermore, it is demonstrated that the learned model (object concept) enables the system to infer properties of unseen objects. The system is evaluated using 49 hand tools in order to show validity of the proposed framework.

1. はじめに

近年，あらゆるシステムの自動化，ロボット化によって，コンピュータによる環境の認識・理解が求められており，様々な研究が行われている．本論文では，その中でも物体の学習や認識に着目する．従来の物体学習や認識は，物体の視覚的特徴のみに基づくものが主流であった．しかし，これは物体の理解という観点から見れば不十分であると考えられる．なぜなら，多くの物体には使用目的や用途が存在しており，これらの情報は視覚的特徴と同様に重要である．つまり，物体はこれらの使用目的や用途を満たすための機能を保有しており，その機能こそが物体の本質であるといえる．特に人が日常的に用いる道具においては使用目的が明確であり，機能の重要性はきわめて高い．また，人が道具を理解し使用するプロセスは，言語のそれと同等といえるほど知能的であるといわれている．そこで本論文では，機械による道具の認識・理解を目的とする．ここでは道具の理解を，経験を通して習得した道具の概念の適用による機能の予測であると考え，道具を機能と視覚情報の関係性によりモデル化することで機械に学習させる．そして，この関係性を通して経験的に道具の認識や機能の推定を行う．このように機能と視覚情報の関係性を導入することは，ギブソンが提唱するアフォーダンス[1)–2)] の計算論的アプローチととらえることができると考えている．

物体の概念をその機能と視覚的情報の関係性と定義した場合，機能をどのように具体的に定義するかを考える必要がある．一般的な道具を使用する際，人は手で道具を持ち，他の物に変化を与える．たとえばハサミで紙を切ったり，ペンで紙に色を塗ったりする．ここでは，この紙などの「変化を与えられるもの」を対象物と呼び，この対象物の視覚的な変化を機能として定義する．対象物の変化は，視覚的特徴として観測することが可能であるため，機械は視覚経験を通してそれ自体範疇的な概念である「機能」を学習し，物体の学習や理解に役立てることが可能となる．

道具からどのような視覚情報を取得するかは，道具の概念の構築におけるもう 1 つの重要な問題である．道具の視覚的特徴を考えた場合，その道具が機能を発揮するために必要な特徴と，そうでない特徴が存在する．前者を機能的視覚特徴，後者を非機能的視覚特徴と呼

†1 電気通信大学電子工学専攻
Department of Electronic Engineering, The University of Electro-Communications
†2 電気通信大学知能機械工学専攻
Department of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications

ぶ．たとえばペンにおいて，先端の紙に触れる尖った部分は機能を発揮するために必要であり，機能的視覚特徴であるといえる．また，装飾のための人形などは非機能的視覚特徴であり，ペンの本質的な機能とは関係がない．このことは，物体の輪郭などの全体的な特徴ではなく，パーツのような局所的な特徴の重要性を示している．そこで本論文では，道具の視覚的特徴として SIFT（Scale Invariant Feature Transform）[3] を用いる．SIFT は事前に計算したコードブックによりベクトル量子化し，最終的に各道具画像における出現頻度として視覚的特徴を表現する．

この視覚的特徴と機能をグラフィカルモデルで表現することで，物体概念を構築する．物体概念の学習とは，機能と視覚情報を利用したカテゴリ分類であるととらえることができる．物体概念の学習後は，機能と視覚的特徴を用いた物体認識だけでなく，未知物体の機能の確率的予測や，必要とする機能から物体を選択することなどが可能となる．また，このモデルを通して局所的な特徴と機能が確率的に結び付くことになり，機能的特徴が観測データから自動的に学習されることを意味している．したがって，提案手法では，機能や機能的形状といった道具を理解するうえで重要な情報を手動で与える必要はなく，これらは機械の視覚経験によって教師なしで自動的に獲得される．

心理学の分野では，人間の教師なしカテゴリ分類に関する研究がなされ，その重要性が指摘されている[4]．教師なしで学習できることで，人は複雑かつ動的な環境に柔軟に適応している．このような能力は，ロボットが我々とともに協調して作業する際にも重要であるといえる．実環境には道具が無数に存在し，それらすべてを人が教えることは現実的ではない．そのため，ロボット自らが，人が道具を使用するシーンを観測することで学習できる能力は非常に重要である．

近年，物体を構成する視覚的特徴パーツに注目し，物体を分類する研究は数多く行われている[5)–10)]．これらは，物体を構成する視覚的特徴パーツに注目して，その物体を構築する視覚的特徴パーツの構成分布により物体の種類ごとの構成比率マップを考えて分類を行っている．ただしこれらの研究では，機能について考えていない．一方，物体を機能で認識することに関する研究も行われている[11)–15)]．これらの研究では機能の辞書や単純な形状テンプレートと機能との関係性を表した辞書などを人手により与え，それを基に物体認識を行っている．しかし，学習の枠組みは取り入れていないため，辞書にない物体に対応できない．したがって新しい物体を認識対象とする場合，それに対応した辞書の拡張を人手で行う必要がある．また文献 16) では，視覚的情報と人間の動作を利用した物体認識に関する研究が行われている．この場合，人間の動作とそれにともなう物体の状況が記述された辞書を用いているため，物体の機能的認識の点では良い成果をあげているが，視覚特徴と機能との関係性を学習することは行っていないため，物体の本質であるといえる機能と視覚特徴の関係性についての情報は得ることができない．また，ヒューマノイドロボットに実際に道具を使用させるための道具の使用モデルを構築する研究も報告されている[17)]．この研究では道具に使用方法をタグ付けすることで汎用性を持たせることが特徴である．ただし，道具のカテゴリ分類は行っておらず，物体（道具）そのものの認識や理解については考えていない．

本論文では道具の分類に確率モデルを用いており，これらの研究とも関連がある．提案モデルでは，機能は変分混合ガウス分布でモデル化され，その学習には変分ベイズ法[18)] を用いる．この手法の利点として，混合数も確率変数として扱い，混合数の確率的な推定が可能な点があげられる．変分混合ガウス分布の混合数を決定する手法として，文献 19) がある．この研究では，混合比の確率分布を考えるのではなく，これをパラメータと見なして点推定を行い，混合係数が 0 に近い要素をモデルから除外することで最適な混合数を決定している．また，文献 20) ではディリクレ過程をパラメータの事前分布とすることで，最適な混合数を決定している．これらの手法では，混合ガウス分布のパラメータを学習するという点では本質的な違いはないため，本論文で提案する機能の学習に適用可能である．しかし，これらの手法自体に多少の違いはあるものの，性能的に大幅な差異はないと思われる．実際，文献 19) では，変分ベイズ法との比較において性能が同等であることを示している．そこで本論文では，混合ガウス分布の構造推定において基本的な手法である，変分ベイズ法を選択する．

また，提案する道具全体のモデルでは，Probabilistic Latent Semantic Analysis（pLSA）[21)] を拡張したモデルを用いて機能と視覚特徴から道具概念を構築している．このモデルに関しても，pLSA をベイズ学習に拡張した Latent Dirichlet Allocation（LDA）[22)] や，文献 20) の Dirichlet Process Mixtures（DPM）が適用可能であると考えられる．本論文では，機能と視覚特徴の関係性をモデル化することが主眼であるため，これらのモデルの中で最もシンプルで直感的に理解しやすい pLSA を採用することとした．

文献 23) では，複数の観測データを同時に分類することが可能なモデルを提案している．この手法では，複数の観測データを分類することで，カテゴリを形成することができる．すなわち，このモデルを適用することで，視覚特徴と対象物の変化から直接道具概念が形成できる可能性がある．しかし本論文では，道具概念だけでなく機能概念も独立した概念として獲得し，基本となる機能概念と視覚情報を基盤としたより上位の物体概念モデルを，pLSA によって表現するという 2 段階の学習手法をとることで学習を容易にする．またこれには，

基本的な機能概念を独立して運用可能であり，人間の意図推定など他の目的にもそのまま適用できるという利点が考えられる．

本論文は，以下次のように構成されている．2 章ではまず，物体の概念モデルについて述べる．3 章では，視覚情報について，4 章では機能のモデル化について述べる．5 章では提案する概念モデルの学習，モデルを用いた認識，推論について述べる．6 章は実験結果であり，7 章で考察を述べ，最後に 8 章で本論文をまとめる．

2. 物体の概念モデル

本論文では物体を「理解」することを，物体の機能の推測が可能であることととらえる．たとえば，シーン（ハサミ）を見ることにより，対象物（紙など）が切断されることを推測できたり機能的視覚特徴を抽出できることが，ハサミに対する理解である．言い換えれば，このような機能と視覚特徴の関係性を学習することが,「ハサミを理解する」ということであると定義する．また物体の名称（単語）は，ある視覚特徴を有しており，使用目的を満たす機能を保有している物体カテゴリに対してラベルとして与えられると考えられる．このことを模式的に表したものが**図 1** (a) である．物体（シーン）が提示されると，そのカテゴリを通して，視覚的特徴量や機能が確率的に出力される．図 1 (a) の理解のモデルをグラフィカルモデルで書き直したものが，図 1 (b) のベイジアンネットワークである．

図における各ノードは確率変数を表しており，O，I，$\boldsymbol{X}_V$，F はそれぞれ，物体の道具カテゴリ，カテゴリ内の各物体に振られた ID（物体 ID），視覚特徴，機能を意味している．これは，文書のトピックモデルである pLSA [21] の拡張であり，文書が物体 I，単語が視覚特徴 $\boldsymbol{X}_V$ と機能 F，文書のトピックが道具カテゴリ O にそれぞれ対応している．すなわち，このモデルでは，確率 $P(O)$ によって物体カテゴリ O が選択され，各条件付き確率に従い物体 I の視覚特徴 $\boldsymbol{X}_V$，機能 F が生成される生成モデルとなっている．

ただしこのモデルは，図 1 (a) を以下の関係を利用して書き換えていることに注意が必要である．

$$P(I)P(O|I)P(\boldsymbol{X}_V|O)P(F|O) = P(O)P(I|O)P(\boldsymbol{X}_V|O)P(F|O) \quad (1)$$

このモデルにおいて直接観測されるノードは，物体 IDI および視覚情報 $\boldsymbol{X}_V$ である．道具カテゴリ O は，視覚情報と機能の関係性に対する範疇的知識であり，直接観測することのできない隠れ変数である．図 1 (b) において，機能ノードは可観測ノードを想定して書かれている．しかし実際には機能も抽象的な範疇的知識であり，直接観測することはできない．前述のように，本論文では機能を対象物の変化として定義する．したがって，実際に観測

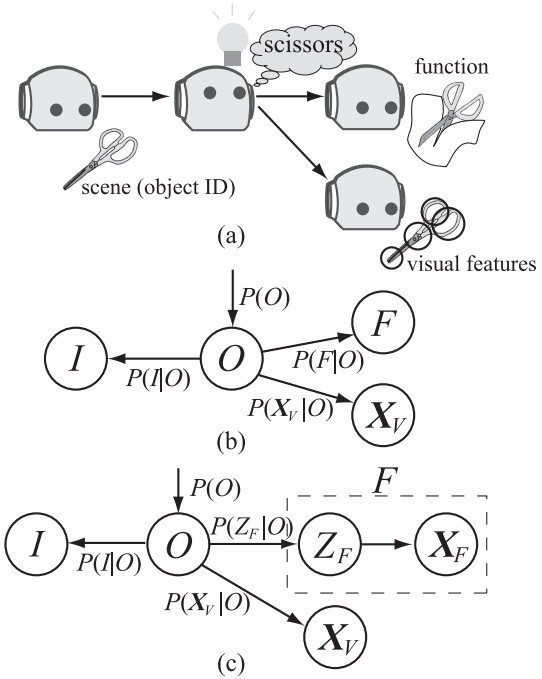

図 1　物体概念のモデル．(a) 概念モデルの模式図，(b) (a) のグラフィカルモデルによる表現，(c) 機能概念を考慮したモデル

Fig. 1　A model of object concept. (a) Schematic diagram. (b) Graphical model representation of (a). (c) A model including abstract function.

されるのは対象物の変化を視覚情報として抽出した特徴ベクトルであり，それらをカテゴライズすることによって得られる機能は非観測である．このことを考慮すると，図 1 (b) は図 1 (c) のように書くことができる．ここで Z_F は，抽象化された機能の概念であり非観測である．一方 $\boldsymbol{X}_F$ は対象物の変化に対する視覚的情報を表しており，可観測である．実際には，図 1 (c) の機能を表す F の部分はガウス混合分布で表現されるが，これについては 4 章で詳しく述べる．

結局，本論文における最終的な問題は，図 1 (c) のモデルにおけるパラメータの推定と，このモデルを用いた推論である．ベイジアンネットワークの特性として，物体のモデルの構築，すなわち学習を行うことにより，不完全データから完全データの推定を行うことが可能

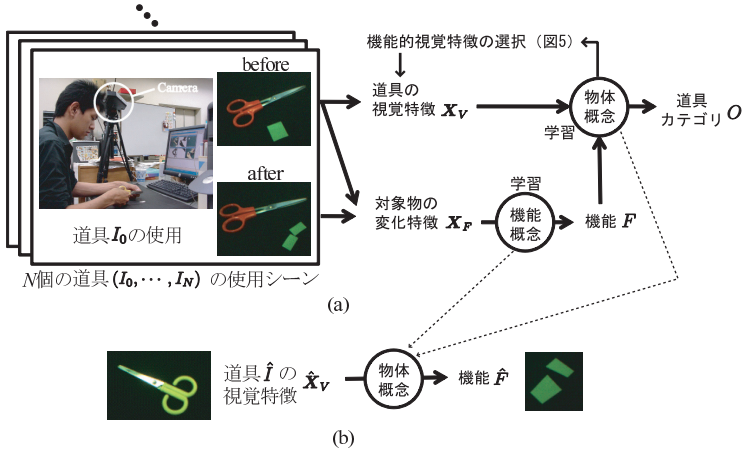

図 2　システム概要．(a) 学習 (b) 機能の推定
Fig. 2　Schematic of the system. (a) The learning pahse. (b) Inference of the function.

となる．よって，学習後は物体の視覚特徴から物体情報や機能情報の推定や，逆に特定の機能を有する物体の視覚特徴を推定することができる．

ここで，機能と道具の学習の流れを**図 2**(a) に示す．まず，ユーザが N 個の道具 $(I_1, \cdots, I_N)$ を使用し，カメラにより動画を取得する．各シーンから，その動きを検出することで道具の使用前と使用後の画像を取得し，道具の使用前の画像から道具の視覚特徴 $\boldsymbol{X}_V$ を，道具使用前と使用後の対象物の変化から対象物体の変化 $\boldsymbol{X}_F$ をそれぞれ計算する．ただし，道具や対象物は黒い背景の中で扱い，それらの自動抽出を容易にした．また，道具と対象物の判定は，それらの位置関係にルール（たとえば道具はつねに左側に置くなど）を決めることですべて自動的に行うこととする．こうして得られた $\boldsymbol{X}_F$ を分類することで，機能概念を学習する．さらに，$\boldsymbol{X}_V$ と機能概念を用いて分類を行うことで，道具（物体）概念の学習を行う．このように学習した道具概念モデルを用いることで，たとえば図 2 (b) のように，未知物体 $\hat{I}$ の視覚特徴 $\hat{\boldsymbol{X}}_V$ から機能 $\hat{F}$ を推定することが可能となる．

3. 物体の視覚特徴

3.1 機能的視覚特徴と非機能的視覚特徴

道具の視覚特徴は 2 つの属性から成り立つと考えられる．1 つは，機能に基づいた特徴（機能的視覚特徴）であり，もう 1 つは，機能に基づかない特徴（非機能的視覚特徴）である．機能的視覚特徴とは，機能を発揮するのに必要となる部分の視覚特徴のことであり，ハサミでは刃の部分や指を入れる部分があり，その位置関係も含むと考えられる．一方，非機能的視覚特徴は，機能とは関係のない部分の視覚特徴であり，ペンについている人形などの装飾を目的にしたパーツなどがあげられる．**図 3** に機能的視覚特徴と非機能的視覚特徴の例を示す．図中の実線が機能的視覚特徴，破線が非機能的視覚特徴である．機能と視覚特徴の関係性を得るためには，機能的視覚特徴のみを扱う必要がある．

図 3　機能的視覚特徴と非機能的視覚特徴の例
Fig. 3　Examples of function related shape and function unrelated one.

3.2 視 覚 特 徴

道具の持つ機能的視覚特徴を考慮すると，道具の視覚特徴として道具の局所的な特徴を取得する必要がある．そこで本論文では，物体の視覚特徴として，SIFT（Scale Invariant Feature Transform）[3] を用いる．SIFT は，画像中の特徴点を抽出し，各特徴点とその周囲の点の方向と勾配を計算し，それらのヒストグラムを特徴ベクトルとして計算する手法である．特徴点は，スケールスペースにより抽出する．SIFT による特徴ベクトルは，シーン全体の明るさや回転の影響を受けにくい．しかし，画像によって抽出される特徴点の数が異なり，道具ごとの視覚特徴量としては扱いにくい．そのためここでは，入力画像から抽出した SIFT を，事前に用意したコードブックによってベクトル量子化し，そのヒストグラムをとることで視覚特徴量とする．この際に用いるコードブックは，様々な道具の画像から SIFT の特徴ベクトルを抽出し，それらの特徴ベクトルを k 平均法でクラスタリングすることで生成する．実際に計算した SIFT 特徴点と，そのヒストグラムの例が**図 4** である．特徴点は画像中の点で示し，ヒストグラムは 500 次元の最初の 10 次元までを掲載した．ハサミでは 5 次元目が，セロテープでは 10 次元目が共通して多く発生していることが分かる．物体概念の学習では，このような共起した特徴と機能の関係を学習することとなる．

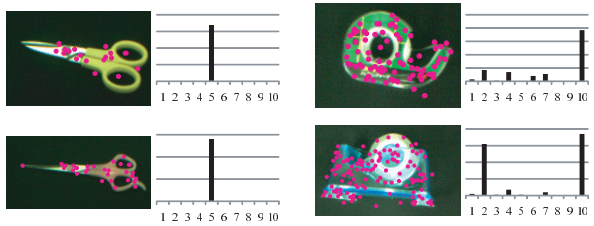

図 4　SIFT 特徴点とヒストグラム
Fig. 4　SIFT keypoints and histograms.

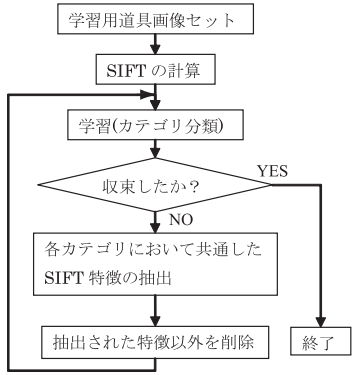

図 5　機能的視覚特徴を考慮した学習アルゴリズム
Fig. 5　Learning algorithm considering functional visual features.

3.3　機能的視覚特徴の学習

SIFT 特徴を用いることで，道具の局所的な特徴を取り出すことができる．しかし，どの局所特徴が機能的特徴であり，どの局所的特徴が非機能的特徴かは判断できない．つまり，機能的視覚特徴は道具によって異なり，これを学習前に事前に与えておくことは不可能である．そこで本論文では，**図 5** に示すアルゴリズムを用いて学習を行うことで機能的視覚特徴を学習することを考える．これは，同じ道具には共通の機能的視覚特徴が多く出現するという仮定を利用したものである．つまり，最初は非機能的特徴を含めたすべての視覚特徴を使って物体の学習を行う．ただし，この際にも同じ道具のカテゴリは互いに共通の視覚特徴を含んでいる可能性が高いこと，また，学習には観測される機能の情報も用いることを考

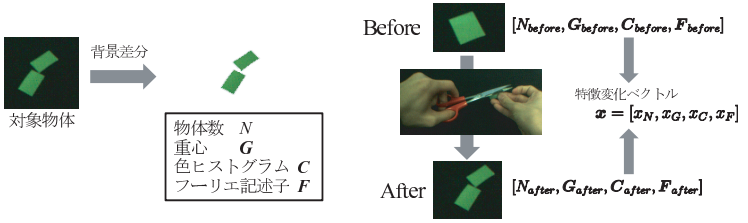

図 6　(a) 対象物体から抽出する特徴量．(b) 対象物体の変化ベクトル
Fig. 6　(a) Feature extraction from a target object. (b) The change vector of a target object.

えると，分類はそれほど正解から外れたものにはならないと考えることができる．したがって，これによって得られるカテゴリごとに共通する視覚特徴を抽出すると，その道具に固有の機能的視覚特徴を抽出することができる．以上の手順を分類が変化しなくなるまで繰り返すことで，最終的にはより正確な道具の分類（学習）と機能的視覚特徴の抽出を行うことができる．

4.　機 能 概 念

4.1　対象物変化の観測

本論文では，機能を考えるうえで対象物に起こる変化に注目する．道具を使用することで対象物に起こる変化は道具使用による効果である．機能はどのような効果を起こすかにより定義されるので，対象物に起こる特徴的な変化を道具の持つ機能としてとらえる[24]．対象物に起こるどのような変化に着目するかは非常に重要であり，これによって観測可能な機能が決定される．ここでは一般的な道具を考慮し，対象物の変化を表すパラメータとして対象物の個数変化 x_N，輪郭変化 x_F，色変化 x_C，重心位置変化 x_G，の 4 つについて考える．また，対象物として以上 4 つの変化を観測しやすい紙を使用した．

図 6 (a) に対象物体から抽出する特徴量について示す．まず，対象物体画像から事前に取得した背景画像を用いた背景差分により物体領域を抽出する．その後，物体領域に対して，物体数 N，重心 $\boldsymbol{G}$，色情報として色ヒストグラム $\boldsymbol{C}$，輪郭情報としてフーリエ記述子 $\boldsymbol{F}$ を計算する．色情報は対象物の色分布を RGB それぞれ 8 次元ずつに量子化して 24 次元のヒストグラムとし，輪郭情報はフーリエ記述子の低周波成分の 10 次元を使用する．

図 6 (b) のように，実際に道具を使うことで，道具の使用前の対象物体の情報

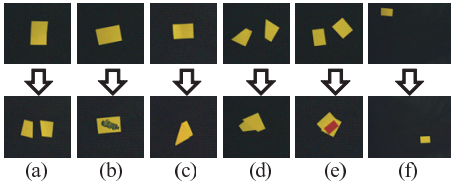

図 7　対象物体へ与えた変化の例．(a) ハサミの使用による変化，(b) ペンの使用による変化，(c) ペンチの使用による変化，(d) ノリの使用による変化，(e) ビニルテープの使用による変化，(f) ピンセットの使用による変化
Fig. 7　Examples of changes in a target object by: (a) scissors, (b) a pen, (c) pliers, (d) a glue, (e) a vinyl tape and (f) tweezers.

$(N_{before}, \boldsymbol{G}_{before}, \boldsymbol{C}_{before}, \boldsymbol{F}_{before})$ と，道具の使用後の対象物体の情報 $(N_{after}, \boldsymbol{G}_{after}, \boldsymbol{C}_{after}, \boldsymbol{F}_{after})$ を求める．その使用する前後の情報から，その道具が及ぼした変化を表す特徴変化ベクトル $\boldsymbol{x} = (x_N, x_G, x_C, x_F)$ を以下の式に従い計算する．ただし，演算中のオーバフローを避けるため，これらの値は 0〜1 の範囲に収まるよう正規化を行う．

(1)　個数変化

$$x_N = N_{after} - N_{before} \tag{2}$$

(2)　重心位置変化

$$x_G = |\boldsymbol{G}_{after} - \boldsymbol{G}_{before}| \tag{3}$$

(3)　色変化

$$x_C = \mathrm{Cr}(\boldsymbol{C}_{after}, \boldsymbol{C}_{before}) \tag{4}$$

(4)　輪郭変化

$$x_F = \mathrm{Cr}(\boldsymbol{F}_{after}, \boldsymbol{F}_{before}) \tag{5}$$

ただし，Cr は正規化相関を表し，$\mathrm{Cr}(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{a} \cdot \boldsymbol{b})/|\boldsymbol{a}||\boldsymbol{b}|$ となる．実際に，道具を使用した際の対象物の変化を**図 7** に示す．

4.2　機能概念の構築

本論文では，観測された対象物の特徴変化ベクトルを混合ガウス分布でモデル化する．パラメータの推定には変分ベイズ法を用いる[25]．変分ベイズ法は，パラメータの推定と同時にモデルの構造も評価することができる．つまり，抽象的な概念である機能の数もデータから推定することが可能となる．混合ガウス分布のグラフィカルモデルは，**図 8** のように書くことができる．これは，図 1 (c) の F の部分に相当する．図 8 において，内側のプレートは学習データ数 N に対する反復計算を表し，$\boldsymbol{x}_n$（$n = 1, \cdots, N$）は可観測の 4 次元の対象物の変化ベクトルを表す．また，$\boldsymbol{\mu}_i$，$\mathbf{V}_i$，α_i はそれぞれ，4 次元混合正規分布の第 i 要素

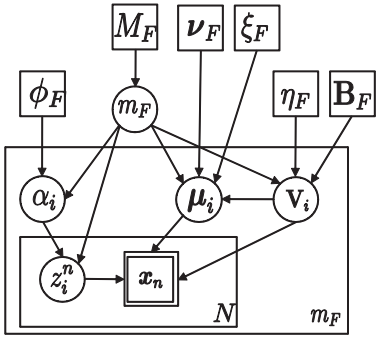

図 8　機能の詳細なグラフィカルモデル
Fig. 8　The detailed graphical model for functions.

（$i = 1, \cdots, m_F$）の平均，精度行列，混合比を表している．m_F はモデルの構造である混合数であり，これは機能がいくつあるかを意味している．これらのパラメータはそれぞれ，事前分布が仮定されている．多項分布 $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_{m_F}\}$ の事前分布は自由度を ϕ_F とするディリクレ分布となり，平均ベクトル $\boldsymbol{\mu}_i$ の事前分布は平均が $\boldsymbol{\nu}_F$，精度行列が $\xi_F \mathbf{V}_i$ となるガウス分布となる．$\mathbf{V}_i$ の事前分布については η_F と $\mathbf{B_F}$ をパラメータとするウィシャート分布となり，確率密度関数は以下のようになる．

$$\mathcal{W}(\mathbf{V}_i | \eta_F, \mathbf{B_F}) \propto |\mathbf{V}_i|^{1/2(\eta_F - d - 1)} \exp\left(-\frac{1}{2}\mathrm{Tr}(\mathbf{V}_i \mathbf{B_F})\right) \tag{6}$$

ただし，$\mathcal{W}()$ はウィシャート分布を表す．また混合数 m_F の事前分布 M_F は，とりうる混合数の最大数を m_{max} としたとき，1 以上 m_{max} 以下の範囲の整数の一様分布を仮定している．z_i^n は，$\boldsymbol{x}_n$ が第 i 要素から生成されているならば 1，さもなくば 0 となる潜在変数であり，これが観測した対象物の変化に対する機能であると考えることができる．

変分ベイズ法では，未知量を周辺化した学習データ集合 $\mathbf{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$ の周辺尤度を考える．

$$\mathcal{L}(\mathbf{D}) = \log P(\mathbf{D}) = \log \sum_{m_F} \sum_{\boldsymbol{Z}_F} \int_{\theta} P(\mathbf{D}, \boldsymbol{Z}_F, \boldsymbol{\theta}, m_F) d\boldsymbol{\theta}, \tag{7}$$

ただし，$\boldsymbol{Z}_F = \{z_i^n\}_{n=1,i=1}^{N,m_F}$ は潜在変数の集合を表し，$\boldsymbol{\theta}$ はモデルパラメータの集合を表している．また，混合数 m_F も確率変数として推定を行うので，すべての混合数に関しても

周辺化を行っている．

ここで，以下のような各未知量ごとに独立性を仮定した，新たな変分事後分布 q を導入する．

$$q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F) = q(m_F) q(\boldsymbol{Z}_F|m_F) \prod_k^K q(\boldsymbol{\theta}_k|m_F) \tag{8}$$

ただし，$\boldsymbol{\theta}$ が独立な K 個のパラメータ $\boldsymbol{\theta}_k$（$k = 1, \cdots, K$）に分解されているものとする．この q を用いて，式 (7) を変形すると以下の式を得る．

$$\begin{aligned}
\mathcal{L}(\mathbf{D}) &= \log \sum_{m_F} \sum_{\boldsymbol{Z}_F} \int_\theta q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F) \frac{P(\mathbf{D}, \boldsymbol{Z}_F, \boldsymbol{\theta}, m_F)}{q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F)} d\boldsymbol{\theta} \\
&= \sum_{m_F} \sum_{\boldsymbol{Z}_F} \int_\theta q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F) \log \frac{q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F)}{P(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F|\mathbf{D})} d\boldsymbol{\theta} \\
&\quad + \sum_{m_F} \sum_{\boldsymbol{Z}_F} \int_\theta q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F) \log \frac{P(\mathbf{D}, \boldsymbol{Z}_F, \boldsymbol{\theta}, m_F)}{q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F)} d\boldsymbol{\theta} \\
&\equiv \mathrm{KL}(q(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F), P(\boldsymbol{Z}_F, \boldsymbol{\theta}, m_F|\mathbf{D})) + \mathcal{F}[q]
\end{aligned} \tag{9}$$

ただし，$\mathcal{F}[q]$ は自由エネルギーと呼ばれ，分布 q を変関数とする汎関数，KL は分布間の距離を表すカルバック・ライブラー距離（KL 距離）である．ここで，$\mathcal{L}(\mathbf{D})$ が変分事後分布 q に依存しないため，$\mathcal{F}[q]$ を最大化することは，q と真の事後分布 P との KL 距離を最少化することと同義である．すなわち，$\mathcal{F}[q]$ を最大化することで，変分事後分布 q は真の事後分布 P の最良の近似となる．

$\mathcal{F}[q]$ を整理すると次式のように変形でき，ラクランジュの未定乗数法を用いて $\boldsymbol{\theta}_k$ について最大化を行うことで，各パラメータの最適事後分布を得ることができる．

$$\begin{aligned}
\mathcal{F}[q] = \sum_{m_F} q(m_F) \Bigg\{ & \left\langle \log \frac{P(\mathbf{D}, \boldsymbol{Z}_F|\boldsymbol{\theta}, m_F)}{q(\boldsymbol{Z}_F|m_F)} \right\rangle_{q(\boldsymbol{Z}_F|m_F), q(\boldsymbol{\theta}|m_F)} \\
& + \sum_{k=1}^K \left\langle \log \frac{P(\boldsymbol{\theta}_k|m_F)}{q(\boldsymbol{\theta}_k|m_F)} \right\rangle_{q(\boldsymbol{\theta}_k|m_F)} + \log \frac{P(m_F)}{q(m_F)} \Bigg\}
\end{aligned} \tag{10}$$

また同様に，$\mathcal{F}[q]$ を $q(m_F)$ に関して最大化することで，混合数の最適変分事後分布 $q(m_F)$ を得ることができる．まず，式 (10) の $q(m_F)$ を含まない項をまとめて $\mathcal{F}_{m_F}$ とおくと次式を得る．

$$\mathcal{F}[q] = \langle \mathcal{F}_{m_F} \rangle_{q(m_F)} + \left\langle \log \frac{P(m_F)}{q(m_F)} \right\rangle_{q(m_F)} \tag{11}$$

$q(\boldsymbol{Z}_F|m_F)$，$q(\boldsymbol{\theta}|m_F)$ は，$\mathcal{F}_{m_F}$ のみに依存するので，$\mathcal{F}[q]$ の $q(\boldsymbol{Z}_F|m_F)$，$q(\boldsymbol{\theta}|m_F)$ に関する最大化は，$\mathcal{F}_{m_F}$ の $q(\boldsymbol{Z}_F|m_F)$，$q(\boldsymbol{\theta}|m_F)$ に関する最大化と等価である．そこで，$\mathcal{F}_{m_F}$ の $q(\boldsymbol{Z}_F|m_F)$，$q(\boldsymbol{\theta}|m_F)$ に関する最大値を $\mathcal{F}^*_{m_F}$ とおくと，混合数の最適変分事後分布は次のようになる．

$$q(m_F) = \frac{P(m_F) \exp\left(\mathcal{F}^*_{m_F}\right)}{\sum_{m_F} P(m_F) \exp\left(\mathcal{F}^*_{m_F}\right)} = C_{m_F} P(m_F) \exp\left(\mathcal{F}^*_{m_F}\right) \tag{12}$$

ただし，C_{m_F} は，$\sum_{m_F} q(m_F) = 1$ とするための規格化定数である．さらに，m_F の事前分布として一様分布 $P(m_F) = M_F$ を仮定しているので，$q(m_F)$ の最大化は，$\mathcal{F}^*_{m_F}$ の最大化と等価である．すなわち，$\mathcal{F}^*_{m_F}$ を $q(\boldsymbol{Z}|m_F)$，$q(\boldsymbol{\theta}|m_F)$ に関して最大化すると同時に，m_F に関しても最大化することにより，最適な混合数 m_F が求まる．

以上が，変分ベイズ法の基本原理となる．最終的に，混合数 m_F の場合の，潜在変数の事後分布は次のようになる．なお，ここでは式の導出を省略するが，詳しくは文献 25) を参照されたい．

$$\begin{aligned}
q(\boldsymbol{Z}_F|m_F) = C \prod_{i=1}^{m_F} \prod_{n=1}^N \exp\Bigg\{ z_i^n \Bigg(& \langle \log \alpha_i \rangle_{q(\boldsymbol{\alpha}|m_F)} + \frac{1}{2} \langle \log |\mathbf{V}_i| \rangle_{q(\mathbf{V}_i|m_F)} \\
& - \frac{1}{2} \mathrm{Tr} \left\{ \langle \mathbf{V}_i \rangle_{q(\mathbf{V}_i|m_F)} \left\langle (\boldsymbol{x}_n - \boldsymbol{\mu}_i)(\boldsymbol{x}_n - \boldsymbol{\mu}_i)^T \right\rangle_{q(\boldsymbol{\mu}_i|m_F)} \right\} \Bigg) \Bigg\}
\end{aligned} \tag{13}$$

また，$\boldsymbol{\alpha}$ の変分事後分布は，$\{\phi_0 + \bar{N}_i\}_{i=1}^{m_F}$ をパラメータとするディリクレ分布となる．

$$q(\boldsymbol{\alpha}|m_F) = \mathcal{D}\left(\{\alpha_i\}_{i=1}^{m_F} | \{\phi_0 + \bar{N}_i\}_{i=1}^{m_F}\right) \tag{14}$$

ただし，$\mathcal{D}()$ はディリクレ分布を表しており，

$$\bar{N}_i = \sum_{n=1}^N \bar{z}_i^n, \qquad \bar{z}_i^n = \langle z_i^n \rangle_{q(z_i^n|m_F)} \tag{15}$$

となる．$\mathbf{V}_i$ の変分事後分布は，$\mathbf{B}_i$ と $\eta_F + \bar{N}_i$ をパラメータとするウィシャート分布に従う．

$$q(\mathbf{V}_i|m_F) = \mathcal{W}(\mathbf{V}_i|\eta_F + \bar{N}_i, \mathbf{B}_i) \tag{16}$$

ただし，$\mathbf{B}_i$ は以下のようになる．

$$\mathbf{B}_i = \mathbf{B}_F + \bar{\boldsymbol{C}}_i + \frac{\bar{N}_i \xi_F}{\bar{N}_i + \xi_F}(\bar{\boldsymbol{x}}_i - \boldsymbol{\nu}_F)(\bar{\boldsymbol{x}}_i - \boldsymbol{\nu}_F)^T, \tag{17}$$

$$\bar{\boldsymbol{x}}_i = \frac{1}{\bar{N}_i}\sum_{n=1}^{N} \bar{z}_i^n \boldsymbol{x}_n, \quad \bar{\boldsymbol{C}}_i = \sum_{n=1}^{N} \bar{z}_i^n (\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_n - \bar{\boldsymbol{x}}_i)^T \tag{18}$$

$\boldsymbol{\mu}_i$ の変分事後分布は，$q(\boldsymbol{\mu}_i, \mathbf{V}_i|m_F)$ を $\mathbf{V}_i$ に関して周辺化することで得ることができ，$\bar{\boldsymbol{\mu}}_i$，$\boldsymbol{\Sigma}\mu_i$，f_{μ_i} をパラメータとする，スチューデントの t 分布となる．

$$q(\boldsymbol{\mu}_i|m_F) = \mathcal{T}(\boldsymbol{\mu}_i|\bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) \tag{19}$$

ただし，$\mathcal{T}()$ は，スチューデントの t 分布を表し，確率密度関数は次式で定義される．

$$\mathcal{T}(\boldsymbol{\mu}_i|\bar{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_{\mu_i}, f_{\mu_i}) \propto \left\{1 + (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)^T (\boldsymbol{\Sigma}_{\mu_i} f_{\mu_i})^{-1} (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}_i)\right\}^{-\frac{d+f_{\mu_i}}{2}} \tag{20}$$

また，各パラメータは次式のようになる．

$$\bar{\boldsymbol{\mu}}_i = \frac{\bar{N}_i \bar{\boldsymbol{x}}_i + \xi_F \boldsymbol{\nu}_F}{\bar{N}_i + \xi_F}, \quad \boldsymbol{\Sigma}_{\mu_i} = \frac{\mathbf{B}_i}{(\bar{N}_i + \xi_F) f_{\mu_i}}, \quad f_{\mu_i} = \eta_F + \bar{N}_i - 3 \tag{21}$$

適当な初期値から始め，潜在変数の事後分布（式 (13)）の更新と，パラメータの変分事後分布の更新（式 (14)，式 (16)，式 (19)）を収束するまで繰り返すことで，$\mathcal{F}[q]$ が局所最大となる変分事後分布が得られる．また，機能の数を表す混合数 m_F も同様に，$\mathcal{F}[q]$ を最大化する m_F を選択する．後に示す実験では，$2 \leq m_F \leq 8$ の範囲でモデル探索を行った．

構築された機能概念モデルを用いてデータ $\boldsymbol{X}_F$ から機能 Z_F^* を決定する際は

$$Z_F^* = \underset{Z_F}{\operatorname{argmax}}\, P(\boldsymbol{X}_F|Z_F) = \underset{j}{\operatorname{argmax}}\, \hat{\alpha}_j \mathcal{N}(\boldsymbol{X}_F|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{V}}_j) \tag{22}$$

とする．ただし，$\mathcal{N}$ はガウス分布を表し，$\hat{\alpha}_j$ は $\boldsymbol{\alpha}$ の最適変分事後分布 $q(\boldsymbol{\alpha}|m_F)$ のモード $\hat{\boldsymbol{\alpha}}$ の第 j 成分．$\hat{\boldsymbol{\mu}}_j$，$\hat{\mathbf{V}}_j$ はそれぞれ，$\boldsymbol{\mu}_j$ の最適変分事後分布 $q(\boldsymbol{\mu}_j|m_F)$ のモードおよび $\mathbf{V}_j$ の最適変分事後分布 $q(\mathbf{V}_j|m_F)$ のモードである．また，物体概念全体の学習や $\boldsymbol{X}_F$ を用いた予測の際は，図 1 (c) における $P(\boldsymbol{X}_F|Z_F)$ に対して，

$$P(\boldsymbol{X}_F|Z_F) = \hat{\alpha}_{Z_F} \mathcal{N}(\boldsymbol{X}_F|\hat{\boldsymbol{\mu}}_{Z_F}, \hat{\mathbf{V}}_{Z_F}) \tag{23}$$

とする．

5. モデル全体の学習と認識・推論への適用

本章では，前述の視覚特徴および機能概念をもとに，物体概念モデル全体の学習を行う手法について述べる．そして，学習したモデルを利用して，物体の認識や未観測情報の確率的推論を行う手法について述べる．

5.1 物体概念の学習

物体学習を始める前に，第 1 段階の学習により構築された機能の概念モデルと SIFT の代表ベクトルを用いて観測データから視覚特徴・機能に関する特徴量ベクトル・尤度を取得する．こうして得られた視覚情報・機能を観測道具情報として物体学習に用いる．ここでの物体学習とは，図 1 (c) のベイジアンネットワークのパラメータ $P(O)$，$P(I|O)$，$P(Z_F|O)$，$P(\boldsymbol{X}_V|O)$ を推定することにほかならない．ただし，すでに述べたように，$P(\boldsymbol{X}_F|Z_F)$ は機能モデルによってすでに決定しているものとする．ここで，O および Z_F は非観測であるため学習には EM アルゴリズムを用いる．今，図 1 のモデルの同時確率は

$$P(I, \boldsymbol{X}_V, \boldsymbol{X}_F, O, Z_F|\boldsymbol{\theta}) = P(O)P(I|O)P(\boldsymbol{X}_V|O)P(Z_F|O)P(\boldsymbol{X}_F|Z_F) \tag{24}$$

となる．したがって，最大化すべき対数尤度は次のようになる．

$$L(D) = \log \sum_{Z_F}\sum_{O} P(I, \boldsymbol{X}_V, \boldsymbol{X}_F, O, Z_F|\boldsymbol{\theta}) \tag{25}$$

ここで上式に Jensen の不等式を適用すると，

$$\begin{aligned} L(D) &= \log \sum_{Z_F}\sum_{O} q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}}) \frac{P(I, \boldsymbol{X}_V, \boldsymbol{X}_F, O, Z_F|\boldsymbol{\theta})}{q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})} \\ &\geq F(q, \boldsymbol{\theta}) = \sum_{Z_F}\sum_{O} q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}}) \log \frac{P(I, \boldsymbol{X}_V, \boldsymbol{X}_F, O, Z_F|\boldsymbol{\theta})}{q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})} \end{aligned} \tag{26}$$

と書くことができる．よって，対数尤度関数を直接最大化するのではなく，下限である $F(q, \boldsymbol{\theta})$ を q と $\boldsymbol{\theta}$ について交互に最大化する．ここで，$\boldsymbol{\theta}$ をモデルのパラメータ，$\hat{\boldsymbol{\theta}}$ をその推定値とする．式 (26) の等号は次式のとき成立する．

$$q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}}) = P(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \boldsymbol{\theta}) \tag{27}$$

したがって，$F(q, \boldsymbol{\theta})$ の q に関する最大化は，

$$q(O, Z_F|I, \boldsymbol{X}_V, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}}) = \frac{P(O)P(I|O)P(\boldsymbol{X}_V|O)P(Z_F|O)P(\boldsymbol{X}_F|Z_F)}{\sum_{Z_F}\sum_O P(O)P(I|O)P(\boldsymbol{X}_V|O)P(Z_F|O)P(\boldsymbol{X}_F|Z_F)} \tag{28}$$

であり，これが E ステップである．一方，$F(q, \boldsymbol{\theta})$ の $\boldsymbol{\theta}$ に関する最大化は，次の Q 関数の最大化となる．

$$Q(\theta) = \langle P(I, \boldsymbol{X}_V, \boldsymbol{X}_F, Z_F, O|\boldsymbol{\theta})\rangle_{q(O,Z_F|I,\boldsymbol{X}_V,\boldsymbol{X}_F,\hat{\boldsymbol{\theta}})} \tag{29}$$

ここで，ラグランジュの未定乗数法を用いて Q 関数を最大化する．最終的に各パラメータの更新式は次のようになる．これが M ステップである．

$$p(O) \propto \sum_I \sum_j \sum_{Z_F} \{n(I, X_{Vj}) q(Z_F, O|I, X_{Vj}, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})\} \tag{30}$$

$$p(I|O) \propto \sum_j \sum_{Z_F} \{n(I, X_{Vj}) q(Z_F, O|I, X_{Vj}, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})\} \tag{31}$$

$$p(X_{Vj}|O) \propto \sum_I \sum_{Z_F} \{n(I, X_{Vj}) q(Z_F, O|I, X_{Vj}, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})\} \tag{32}$$

$$p(Z_F|O) \propto \sum_j \sum_I \{n(I, X_{Vj}) q(Z_F, O|I, X_{Vj}, \boldsymbol{X}_F, \hat{\boldsymbol{\theta}})\} \tag{33}$$

ただし，X_{Vj} は，$\boldsymbol{X}_V$ の j 次元目を表し，$n(I, X_{Vj})$ は物体 I の視覚特徴 X_{Vj} の生起回数である．

ここで示した EM アルゴリズムは，図 5 における「学習（カテゴリ分類）」の部分であり，実際の学習では，機能的視覚特徴を選択しながらカテゴリが変化しなくなるまで繰り返されることになる．

5.2　物体の認識と推論

物体の概念モデルを構築した後，そのパラメータと観測情報を用いて，未観測情報の確率的推論を行うことができる．まず，物体の視覚情報（視覚特徴）$\boldsymbol{X}_V$ および機能に関する情報（対象物変化）$\boldsymbol{X}_F$ が観測された場合の物体（カテゴリ）認識について考える．この際まず，機能モデル（式 (23)）に基づいて各機能に対する尤度を計算する．そして，次式より物体の認識を行うことができる．

$$\begin{aligned}&\operatorname*{argmax}_O P(O|\boldsymbol{X}_V, \boldsymbol{X}_F, I)\\ &= \operatorname*{argmax}_O \frac{P(O)P(I|O)P(\boldsymbol{X}_V|O)\sum_{Z_F}\{P(Z_F|O)P(\boldsymbol{X}_F|Z_F)\}}{\sum_O[P(O)P(I|O)P(\boldsymbol{X}_V|O)\sum_{Z_F}\{P(Z_F|O)P(\boldsymbol{X}_F|Z_F)\}]}\end{aligned} \tag{34}$$

ただし，この認識は学習済みの物体 I に対する認識である．提案したモデルで未学習の物体 $\hat{I}$ を扱うために，新たな物体に対して $P(\hat{I}|O)$ および $P(O)$ を再計算する．この再計算には，前述の EM アルゴリズムを用いるが，その際に $P(\boldsymbol{X}_V|O)$，$P(Z_F|O)$ は更新しない．また，機能モデルも固定する．このようなヒューリスティックスは，文献 21) で pLSA に対して提案されたもので，fold-in ヒューリステックスと呼ばれる．

次に，物体 $\hat{I}$ の視覚特徴 $\boldsymbol{X}_V$ のみが観測された場合に，機能を推定することを考える．この際の機能推定は，次のように行うことが可能である．

$$\operatorname*{argmax}_{Z_F} P(Z_F|\boldsymbol{X}_V, \hat{I}) = \operatorname*{argmax}_{Z_F} \frac{\sum_O P(O)P(\hat{I}|O)P(\boldsymbol{X}_V|O)P(Z_F|O)}{\sum_O P(O)P(\hat{I}|O)P(\boldsymbol{X}_V|O)} \tag{35}$$

ただし，この際にも上述の fold-in ヒューリステックスを利用する．これ以外の組合せに関しても，同様に確率的な推論を行うことが可能である．

6. 実　　験

実験に用いた道具を**図 9** に，各種類の道具数，観測データ数を**表 1** に示す．実験では計 49 個の道具を，A セット 33 個および B セット 16 個に分割し，特に断りのない限り A セットを学習，B セットをテストに用いることにする．また，表中の T1〜T8 は正解となる物体カテゴリを表している．

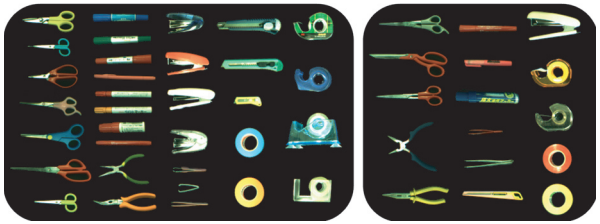

図 9　実験に用いた道具
Fig. 9　The tools used in the experiment.

表 1　実験に用いた道具数
Table 1　Number of tools in the experiment.

物体カテゴリ	番号	A set	B set	合計
ハサミ	T1	7	3	10
ペン	T2	8	3	11
ペンチ	T3	2	2	4
ピンセット	T4	3	2	5
カッタ	T5	3	1	4
ホッチキス	T6	4	1	5
セロテープ	T7	4	2	6
ビニールテープ	T8	2	2	4

実験では，実際に道具を使用し，動き検出によって検出した動き開始直前と動き終了直後の画像を，それぞれ道具使用前・使用後の画像として用いる．各画像に対して，あらかじめ取得した背景画像を用いた背景差分により，道具と対象物の領域を抽出する．また，道具は必ず左側に置くというルールを決め，道具と対象物の判定を行う．使用前の道具領域から道具の視覚特徴を計算し，道具使用前後の対象物領域から 4.1 節で述べた手法で特徴変化ベクトルを計算する．以上の処理を各道具につき 10 回ずつ行うことで，計 490 個のデータを構築し，実験に用いる．

6.1　機能概念の構築

A セットの道具を使用して観測された対象物の特徴変化ベクトルから，機能の学習を行った．混合数 m_F は 2 から 8 へ変化させ，それぞれの場合の自由エネルギーを計算した．その結果が**図 10** であり，混合数 $m_F = 6$ のときの自由エネルギーが最大となった．これは機能が 6 つあることを意味している．このときの機能の分類結果が，**図 11** (a) である．この図において，縦軸が物体カテゴリを，横軸が分類された機能を表しており，分類された個数を輝度の大きさで表した．すなわち輝度値が高いところほど，多くの物体がその機能へ分類されたことを意味している．

この結果から，Function ID 1 はハサミ（T1）とカッタ（T5）が含まれており,「切断」の機能として分類されていることが分かる．同様に，Function ID 2〜5 はそれぞれ「色変化」「変形」「移動」「接着」として分類されている．さらに，接着することで色変化を引き起こすビニールテープ（T8）は,「接着」や「色変化」として分類されることなく，正しく 1 つの機能（接着&色変化）として分類されている．

実際に学習された各物体における特徴変化ベクトルの平均 $\boldsymbol{\mu}_i$（正確には，$q(\boldsymbol{\mu}_i|m_F = 6)$ のモード）を**図 12** に示した．x_N は個数の増減であり，正規化されているため，0.5 付近

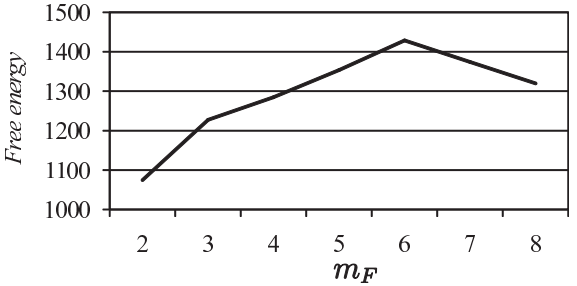

図 10　機能数 m_F と自由エネルギー $F(q)$
Fig. 10　Number of functions m_F versus free energy $F(q)$.

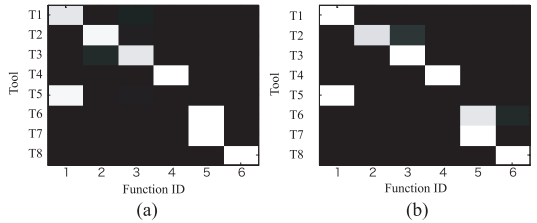

図 11　(a) 機能の学習結果，(b) 機能の認識結果
Fig. 11　(a) The result of function learning, (b) The result of function recognition.

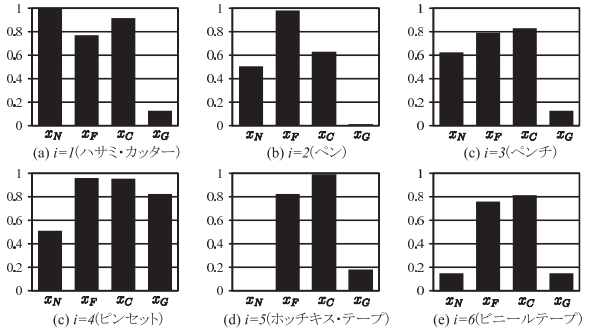

図 12　学習された機能特徴ベクトル
Fig. 12　Learned functional feature vectors.

が個数変化なし，$x_N > 0.5$ で個数の増加，$x_N < 0.5$ で個数の減少を表している．(a) ハサミ・カッタでは個数が増加し，(e) ホッチキス・テープや (f) ビニルテープでは個数が減少していることが分かる．x_F は，使用前後での形状の相関であり，値が小さいほど形状の変化が大きいことを表している．(b) ペンや (d) ピンセットは，形に変化を及ぼさないため，他に比べて高い値となった．x_C は使用前後での色の相関であり，値が小さいほど色変化が大きいことを意味している．(b) ペンで色変化が最も大きく，(c) ペンチや (f) ビニルテープでも色が変化していることが分かる．(c) ペンチで色変化が生じたのは，対象物である紙を変形させることで影が生じたのが原因であると考えられる．また，x_G は重心位置の変化であり，値が大きいほど位置が変化していることを表しており，(d) ピンセットにおいて，位置が大きく変化している．このように，道具の機能がおおむね正しく学習できていることが分かる．

次に学習した機能モデルを用いて，学習に用いていない B セットの道具を使用した際の特徴変化ベクトルに対して，機能認識を行った．その結果が図 11 (b) である．ここでは，分類（認識）精度を以下のように定義し，その値を求めた．

$$Acc = \frac{1}{8}\sum_{t=T1}^{T8}\frac{(\text{道具 } t \text{ のうち正しいカテゴリに分類された数})}{(\text{道具 } t \text{ の総数})} \times 100 \qquad (36)$$

その結果，学習（A セット）96.6%，認識（B セット）97.1%と，学習・認識ともに 9 割以上の精度であった．

6.2　物体概念の構築

前節の実験で得られた機能の学習結果と，物体の視覚特徴から物体概念の構築を行った．さらに，機能の情報を用いずに視覚特徴のみで物体概念の構築を行った．それらの結果が**図 13** である．縦軸が正解となる物体カテゴリを横軸が分類されたカテゴリの ID を表しており，道具が各カテゴリに分類された数を輝度の大きさで表している．すなわち，図 13 (a) が，すべての道具が正しく分類されていることを表している．視覚特徴のみによる分類（図 13 (b)）では，セロテープ（T7）とビニルテープ（T8）は似た視覚特徴を持っているため，同じカテゴリとして分類されてしまっている．また，他の道具もノイズの影響により複数のカテゴリに分かれてしまっている．一方，機能情報と視覚情報を用いた分類では，視覚情報のみの分類に比べ正しく分類できていることが分かる．分類の精度は，視覚情報のみの場合が 63.1%，視覚情報と機能情報を用いた場合が 93.0%となり，精度が約 30%向上している．視覚情報だけでは分類できない物体であっても，機能の情報を手がかりとすることで正しく分類することが可能となる．

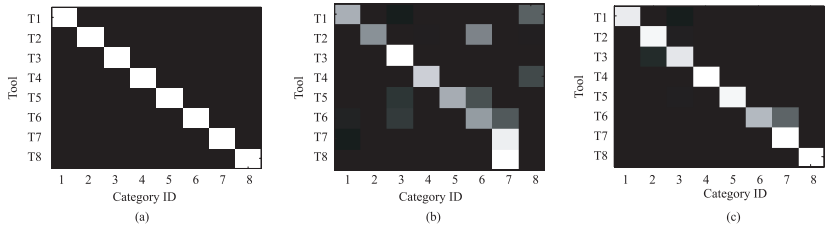

図 13　(a) 正解となる分類，(b) 視覚特徴のみを用いた分類，(c) 機能と視覚特徴を用いた分類
Fig. 13　(a) The correct categorization, (b) The categorization with visual information, (c) The categorization with visual and function information.

6.3　機能的視覚特徴

次に，図 5 に示した機能的視覚特徴を考慮したアルゴリズムの有効性を検証する実験を行った．提案アルゴリズムでは，ランダムな初期値から始めて，分類の変化がなくなるまで図 5 のループを繰り返すことになる．本実験では，1,000 個のランダムな初期値から 1,000 個のモデルを学習し，各モデルに対して分類の精度を求め評価を行った．以下，i 個目のモデルにおいて，提案アルゴリズムのループを l 回繰り返した際の分類の精度を Acc_{il} と表記する．

提案アルゴリズムの繰返しによる精度の変化が**図 14** (a) である．横軸が繰返し回数 l，縦軸が各 l におけるモデル 1,000 個分の精度の平均（$(\sum_{i=1}^{1000} Acc_{il})/1000$）である．提案アルゴリズムを繰返すことで，精度が増加し正しい分類に近づいている．これは，繰返しにより

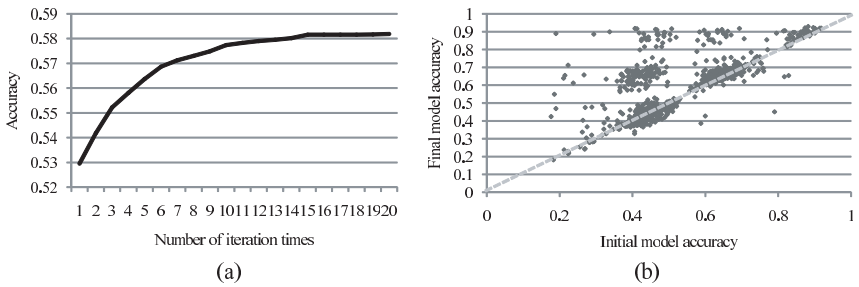

図 14　(a) 精度の平均と繰り返し回数，(b) 初期モデルと最終モデルの精度
Fig. 14　(a) Number of iterations versus accuracy, (b) Accuracy of initial models versus that of final models.

非機能的視覚特徴が除去され，機能的視覚特徴が正しく抽出されているためであると考えられる．

次に，i 個目のモデルの，すべての視覚特徴を用いたモデルと，そのモデルを初期モデルとして，提案アルゴリズムを 20 回繰り返した最終モデルの精度（Acc_{i20}）の関係を図 14 (b) に示した．

横軸が，すべての視覚特徴を用いた初期モデルの精度，縦軸が最終モデルの精度となっており，図中の破線が精度の改善がなく，初期モデルの精度と最終モデルの精度が等しくなる境界である．すなわち，この直線より上にあるモデルでは，精度が向上したことを表しており，提案アルゴリズムによって精度が向上する傾向にあることが分かる．さらに，初期モデルでは精度が低い局所解に陥った場合であっても，最終的には 8 割以上の精度となるモデルも存在している．このように，すべての視覚特徴を用いて学習を行うよりも，その道具特有の機能的視覚特徴を抽出することで，より分類の精度が向上するといえる．

6.4　認 識 実 験

前節の学習において最も尤度が高くなったモデルを用いて，学習には使用していない B セットの道具の認識実験を行った．まず視覚情報と機能情報の 2 つを用いた物体カテゴリの認識を行った．その結果が**図 15** (a) である．この認識の精度は 92.9%となり，学習したモデルを用いることで，未知の物体でも正しく認識できることが分かる．

次に視覚情報のみからその道具の機能の認識を行った．その結果が図 15 (b) である．認識の精度は 95.4%となり，視覚情報のみから機能情報を正しく予測できていることが分かる．これは，視覚情報と機能情報とが関係性を持っていることを示しており，その関係性が正しく学習されていることを意味している．

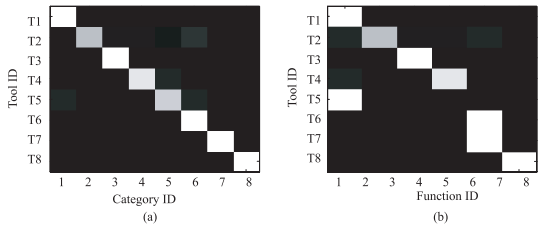

図 15　認識実験結果．(a) 視覚情報と機能情報からカテゴリの認識，(b) 視覚情報のみから機能の認識
Fig. 15　The results of recognition: (a) Recognition of categories from visual information and functions, (b) Recognition of funcitons only from visual information.

6.5　機能に関連した視覚特徴

学習したモデルを用いることで，視覚情報のみから正しく機能情報の予測が可能であることを示した．これは，特定の視覚特徴が特定の機能と結び付いていることを意味している．実際にこのような視覚特徴を図示した．特定の機能 Z_F と強く結び付いている視覚特徴 $\bar{X}_{VF}$ は，以下のように求めた．

$$\bar{X}_{VF} = \underset{X_V}{\operatorname{argmax}}\, P(Z_F|X_V) \tag{37}$$

この式より得られた機能の視覚特徴を**図 16** に図示した．はさみでは切断の特徴，ペンでは色変化の特徴，ペンチでは変形の特徴が現れている．すなわち，道具特有の機能的視覚特徴が正しく学習されていることが分かる．

7.　考　　察

7.1　特徴量について

今回の実験では，道具の視覚特徴・機能の特徴変化ベクトルは，単一の黒い背景において

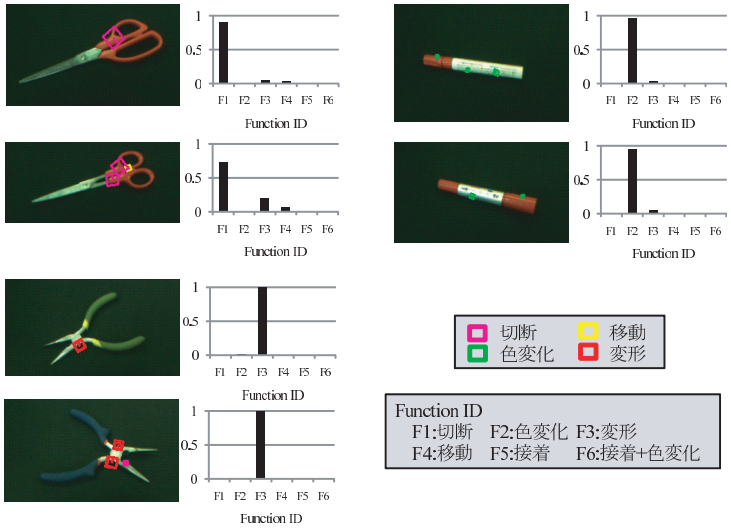

図 16　機能的視覚特徴と機能の認識結果
Fig. 16　Functional visual features and results of function recognition.

取得している．実環境において背景の単一性は保障されないため，実環境で動作させる場合には，この点を改良しなければならない．ただし，道具の視覚特徴に関しては，複雑な背景に対してもある程度の頑健性を持つ SIFT 特徴量を使用しているため，この問題を解決できる可能性を持っている．一方で，今回用いた機能の特徴変化ベクトルは，このような頑健性を持たない．対象物体の変化として，形・色・位置といった大局的な特徴を考えており，視覚特徴に用いたような局所的な特徴を使用することが難しかったためである．今後，実環境における対象物体の特徴変化を取得するために，複雑な背景においても頑健な特徴量を導入しなければならないと考えられる．

7.2　モデルの妥当性について

本論文では，基本となる機能概念を学習し，機能と視覚特徴から道具概念全体の構築を行っている．このように 2 段階のモデル化を行うことには，いくつかの利点がある．まず，基本概念である機能を学習し抽象化することで，上位の道具概念の学習を容易にしている．さらに，新たな基本概念（使い方など）の導入が容易であるといった点も利点としてあげられる．基本概念のモデル化が独立で行われることで，新たな基本概念の導入の際に変更する必要がなく，それらの依存関係が上位の物体概念で獲得されることになる．すなわち，上位の概念を比較的シンプルなモデルである pLSA でモデル化することによって，拡張が容易になることが期待できる．しかし，pLSA は点推定であり，一般的にベイズ推定に比べ性能が低い点が指摘されている[22]．さらに，pLSA ではカテゴリ数を決定することができないといった欠点もある．こうした問題点は，pLSA を DPM などに置き換えることで改善できる可能性があるが，この点に関しては，今後の課題である．

7.3　アフォーダンスとの関連性

アフォーダンスとは，知覚と行為に関する理論であり，情報は環境にある，つまり環境が行為をアフォードすることが，動物の活動の原点となっているという考え方である．近年では，アフォーダンスの考え方がさらに拡張され，たとえばデザインの分野では，アフォーダンスを事物の知覚された特徴あるいは現実の特徴，とりわけ，そのものをどのように使うことができるかを決定する最も基礎的な特徴として定義している[2]．また，道具自体が作り手の意図した機能や使い方をアフォードしていると考えることができ，これを文化的アフォーダンスの一部と考える場合もある．

実験において，提案モデルにより，視覚的機能特徴の抽出が可能となった（図 16）．このような機能と関係した特徴を認識できることは，文献 2) のノーマンが提唱するアフォーダンスの概念である．ギブソンのオリジナルでは，アフォーダンスを客観的な事実としてとらえているの対して，ノーマンは，経験に基づいた主観も含んでいる．図 15 (b)，図 16 の結果は，このノーマンのアフォーダンスを示していると考えられる．つまりシステムは，特定の視覚特徴量から，その機能をアフォードされることになる．重要なことは，これがシステムの視覚経験によって形成された物体概念モデルを通して行われることである．

文献 2) ではさらに，機能の手がかりとして制約が重要であるとしている．たとえば，ハサミの穴は何かを入れることをアフォードすると同時に，その大きさが制約となり，指を入れて使用することを示していることになる．本論文で提案したモデルは，こうした身体的な制約や使い方は考慮されていないため，この点においては道具の概念モデルとして不十分であるといえる．この点に関しては，ロボットを用いることで身体性を利用すること，また使い方をモデルに組み込むことが考えられ，これらは今後の課題である．ただし，使い方のモデル化は，システムが人の道具使用を観察し，その動きを解析することである程度は可能であると思われる．著者らは，その際に必ずしも身体が必要であるとは考えていない．もちろん身体がなければ，道具を実際に使うことはできない．しかし，使い方やその結果を想像することは可能であり，これはミラーニューロンの働きと同様，他人の行動の予測や意図理解を可能にする．この際，システムには人間と同様の身体性が暗に仮定されることとなる．また，道具使用時の動きを解析するためには，人間の身体に関するモデルを持たなければならない．

トマセロは，子供が他人の道具使用から，その道具の意図の知覚を学習すると述べており，これをインテンショナルアフォーダンスと名付けた[27]．提案したモデルは，インテンショナルアフォーダンスの学習をモデル化していると考えることもできる．しかし本論文では，意図や目的を考慮しているわけではなく，道具の機能を考えているだけである．また，本論文における機能の定義は，機能の 2 次的な側面に過ぎず，かなり限定されたものである．これは，機能自体が非観測な情報であり，これを推定する手がかりとして可観測な視覚情報を用いているためである．したがって，目に見える十分な変化を起こさない物体の機能を知覚することはできない．文献 28) では，動物のアフォーダンス学習に対して，次のように指摘している．『動物は，他の動物の行動を観測することでその行動そのものでなく，その行動が及ぼした環境変化について学習する』．これは，提案するモデルの基礎的なレベルでの妥当性を示していると考えている．しかし，提案する機能モデルが，より複雑な機能をモデル化するのに十分でないことは明らかである．より複雑な機能を知覚し学習するために，単純にモダリティや特徴量を増やすことが考えられるが，依然として意図や目的といった非常に複雑な概念をモデル化するためにどのような特徴を用いるかは難しい問題であり，

今後の課題である．

8. おわりに

本論文では，機能と視覚特徴に着目した物体の学習・認識の手法を提案した．機能を対象物の変化を観測しモデル化することで人手で機能を定義したり，拡張することなく，機能の概念を構築することが可能となった．また，実験から物体の視覚特徴と機能という多角的な情報で物体を学習することで，視覚特徴のみの場合よりも精度が向上するという結果が得られ，物体学習・認識における機能の有効性を示すことができた．さらに提案モデルを利用することで，道具の見た目からその機能を予測することが可能であり，限られた形ではあるが本論文で定義した道具の理解を実現したといえる．このようなモデルは，実際に人の作業をサポートをする場合にも有効であると考えられる．たとえば，人が紙を切りたい場合に，ロボットは物を切る機能を持ち合わせた道具の視覚特徴を推定することで，物を切ることができそうな道具を選んで人に渡すことができる．また，人がハサミを取るよう指示したがハサミがない場合に，同じ機能を保有しているカッタを渡すことが可能である．

今後の課題として，実際に道具を使用するシーンの観測があげられる．使用中のシーンを観測することで，道具の各パーツの機能（切る機能，手で把持される機能など）を考えることができ，道具の理解・認識についてさらに追求できる可能性がある[26]．また，道具の使い方をモデル化できれば，機械が自律的に道具を使えるようになり，さらに，1つの道具でも使い方によって複数の機能を発揮しうることを理解可能となる．使い方と機能の関係性を学習することで，このような高度な道具の理解が実現できる可能性があり，今後ロボットを用いることで，こうした身体性に基づく道具の理解と使用を実現させたいと考えている．

謝辞　本研究は，科研費（20500186，20500179）および新学術領域研究「伝達創成機構」の助成を受け実施したものである．

参考文献

1) Gibson, J.J.: *The ecological approach to visual perception*, Lawrence Eribaum, Hillsdale, NJ (1979).
2) D.A. ノーマン（著），野島久男（訳）：誰のためのデザイン？—認知科学者のデザイン原論，新曜社 (1990).
3) Lowe, G.D.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2 (2004).
4) Heit, E., Brockdorff, N. and Lamberts, K.: Categorization, Recognition and Unsupervised Learning, *Memory and Mind*, Gluck, A.M., Anderson, R.J. and Kosslyn, M.S. (Eds.), pp.325–342, Erlbaum (2007).
5) Fergus, R., Perona, P. and Zisserman, A.: Object class recognition by unsupervised scale-invariant learning, *Proc. CVPR*, Vol.2, pp.264–271 (2003).
6) Sivic, J., Russell, C.B., Efros, A.A., Zisserman, A. and Freeman, T.W.: Discovering object categories in image collections, AI Memo, 2005-005:1–12 (2005).
7) Sivic, J., Russell, C.B., Efros, A.A., Zisserman, A. and Freeman, T.W.: Discovering objects and their location in images, *Proc. ICCV2005*, Vol.1, pp.370–377 (2005).
8) Fergus, R., Perona, R. and Zisserman, A.: Using Multiple Segmentations to Discover Objects and Their Extent in Image Collections, *Proc. CVPR2006*, Vol.2, pp.1605–1614 (2006).
9) Fei-Fei, L. and Perona, P.: A Bayesian Hierarchical Model for Learning Natural Scene Categories, *Proc. CVPR2005*, Vol.2, pp.524–531 (2005).
10) Fei-Fei, L., Fergus, R. and Perona, P.: One-Shot Learning of Object Categories, *IEEE Trans. PAMI*, Vol.28, No.4, pp.594–611 (2006).
11) Stark, L., Bowyer, K., Hoover, A. and Goldgof, D.B.: Recognizing object function through reasoning about partial shape descriptions and dynamic physical properties, *Proc. IEEE*, Vol.84, No.11, pp.1640–1656 (1996).
12) Rivlin, E., Dickinson, S.J. and Rosenfeld, A.: Recognition by functional parts, *Computer Vision and Image Understanding*, Vol.62, No.2, pp.164–176 (1995).
13) Woods, K., Cook, D., Hall, L., Bowyer, K. and Stark, L.: Learning membership functions in a function-based object recognition system, *Journal of Artificial Intelligence Research*, Vol.3, pp.187–222 (1995).
14) 服部洋一，黄瀬浩一，北橋忠宏，福永邦雄：動的機能のモデルに基づく物体の機能認識，情報処理学会論文誌，Vol.36, No.10, pp.2277–2285 (1995).
15) Shimshoni, I., Rivlin, E. and Soldea, O.: Efficient Search and Verification for Function Based Classification from Real Range Images, *CVIU*, Vol.105, No.3, pp.200–217 (2007).
16) Kojima, A., Higuchi, M., Kitahashi, T. and Fukunaga, K.: Toward a cooperative recognition of human behaviors and related objects, *Proc. 2nd International Workshop on Man-Machine Symbiotic Systems*, pp.195–206 (2004).
17) 小倉　崇，岡田　慧，稲葉雅幸：注目点を持つ幾何モデルを利用したヒューマノイドの道具利用動作の生成法，第23回日本ロボット学会学術講演会公演論文集，1F15 (2005).
18) Attias, H.: Inferring Parameters and Structure of Latent Variable Models by Variational Bayes, *Proc. 15th Conference on Uncertainty in Artificial Intelligence*, pp.21–30 (1999).
19) Corduneanu, A. and Bishop, M.C.: Variational Bayesian Model Selection for Mixture Distributions, *Proc. International Conference on Artificial Intelligence and*

Statistics, pp.27–34 (2001).
20) Blei, M.D. and Jordany, I.M.: Variational Inference for Dirichlet Process Mixtures, *Journal of Bayesian Analysis*, Vol.1, No.1, pp.121–144 (2006).
21) Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis, *Machine Learning*, Vol.42, pp.177–196 (2001).
22) Blei, M.D., Ng, Y.A. and Jordan, I.M.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol.3, pp.993–1022 (2003).
23) Barnard, K., Duygulu, P., Forsyth, D., Freitas, N., Blei, D. and Jordan, I.M.: Matching Words and Pictures, *Journal of Machine Learning Research*, Vol.3, pp.1107–1135 (2003).
24) Shinchi, Y., Sato, Y. and Nagai, T.: Bayesian Network Model for Object Concept, *Proc. ICASSP2007*, Vol.2, pp.473–476 (2007).
25) 上田修功：ベイズ学習：変分ベイズ学習の応用例，電子情報通信学会誌，Vol.85, No.8, pp.633–638 (2002).
26) 中村友昭，新地康人，長井隆行：グラフィカルモデルを用いた物体概念モデル，FIT2008, RF-002 (2008).
27) Tomasello, M.: *The Cultural Origins of Human Cognition*, Harvard University Press (1999).
28) Tomasello, M.: Do Apes Ape?, *Social Learning in Animals: The roots of culture*, Heyes, C.M. and Galef, B.G. (Eds.), pp.319–346, Academic Press (1996).

(平成 21 年 9 月 19 日受付)
(平成 22 年 5 月 6 日採録)

中村　友昭

平成 21 年電気通信大学大学院電気通信学研究科修士課程修了．現在，同大学院博士課程在学中．知能ロボットに関する研究に従事．

長井　隆行（正会員）

平成 5 年慶應義塾大学理工学部卒業．平成 9 年同大学大学院博士課程修了．博士（工学）．平成 10 年電気通信大学電子工学科助手．平成 15 年カリフォルニア大学サンディエゴ校客員研究員．平成 16 年電気通信大学大学院電気通信学研究科助教授．平成 22 年同大学院情報理工学研究科准教授（現職）．マルチメディア信号処理，知能システム，知能ロボティクスに関する研究に従事．