1J1-OS13-3

複数のマルチモーダルLDAを用いた抽象的概念の形成

Forming Abstract Concept Using Multiple Multimodal LDA Models

中村 友昭[*1] 長井 隆行[*2] 岩橋 直人[*3]
Tomoaki NAKAMURA Takayuki NAGAI Naoto IWAHASHI

[*1]電気通信大学電子工学専攻
Dept. of Electronic Engineering, The University of Electro-Communications

[*2]電気通信大学知能機械工学専攻
Dept. of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications

[*3]情報通信研究機構
National Institute of Information and Communications Technology

In this paper we propose an LDA-based framework for multimodal categorization. The robot uses its physical embodiment to grasp and observe an object as well as listen to the sound during the observing period. This multimodal information is used for categorizing and forming multimodal concepts. The main issue, which is tackled in this paper, is granularity of categories. The categories are not fixed but varied according to context. Selective attention is the key to model this granularity of categories. For example, the category (concept) "soft" is formed by paying attention to haptic modality, while the category "maraca" is based on audio-visual information. These facts motivate us to use a set of weights to the perceptual information. Obviously, as the weights change, the categories vary. In the proposed model, various sets of weights and model structures are assumed. Then the LDA-based categorization is carried out many times that results in a variety of models. In order to make the categories (concepts) useful for inference, significant models should be selected. The selection process is carried out through the interaction between the robot and the user. These selected models enable the robot to infer unobserved properties of the object. For example, the robot can infer audio information only from its appearance. Furthermore, the robot can describe appearance of any objects using some suitable words, thanks to the connection between words and perceptual information. The proposed algorithm is implemented on a robot platform and some experiments are carried out to validate the proposed algorithm.

1. はじめに

事物のカテゴリ分類は，人間の知的活動において重要な役割を果たしている．人間はカテゴリを形成することにより，カテゴリを通した未観測情報の予測を可能とし，これが事物の理解の基礎となっている．すなわちロボットにおいても，このようなカテゴリ分類する能力が非常に重要であると考えられる．

そこで著者らは，文献 [中村 08] において，pLSA (probabilistic Latent Semantic Analysis)[Hofman 01] 及び LDA (Latent Dirichlet Allocation)[Blei 03] を拡張したマルチモーダルカテゴリゼーションを提案した．この研究では，複数のモダリティを用いることにより，より人間の感覚に近いカテゴリを形成することが可能となることを示した．さらに，提案手法は確率モデルに基づいており，学習したグラフィカルモデルを用いることで，未学習物体のカテゴリ認識が可能である．また，学習したモデルを用いた未観測モダリティ情報の推定を可能とした．例えば，物体を見ることで得られる視覚情報から，その物体の硬さやどのような音がするかといった情報を確率的に推定することが可能である．これはまさに人間が日々行っている物体のカテゴリを通した機能の予測であり，提案手法によりその機能をロボットに実装することが可能となった．

しかし，pLSA や LDA はカテゴリ数をあらかじめ人手で与えなければならないという欠点があり，想定した粒度のカテゴリのみしか獲得することができず，文献 [中村 08] では，物体カテゴリの形成のみを目的とした．しかし，人が用いているカテゴリは，抽象的なものから具体的なものまで様々な粒度で存在している．さらに，抽象的なカテゴリには，色カテゴリや触覚カテゴリのように特定のモダリティと結びついたものが存在する．そこで，本稿ではマルチモーダル LDA を拡張し，このような特定のモダリティとの結びつきの強さを考慮した，様々な粒度のカテゴリ分類を行うことで，物体概念も含めた様々な粒度の概念の獲得を目指す．まず，図 1 の破線矩形内のように，物体から得られたマルチモーダル情報から，モダリティとの結びつきや，カテゴリ分類の粒度を様々に変化させた複数のマルチモーダル LDA の学習を行う． しかし，様々な粒度のカテゴリを形成すると，その中には人の感覚には即さないカテゴリも多く形成されることになる．そのため提案手法では，人との対話を通して，形成されたカテゴリと単語を結びつけ，人の感覚に即したカテゴリの選択を行う (図 1 実線矩形内)．最終的に，ロボットはマルチモーダル情報により形成されるカテゴリと，それを表す単語，さらにカテゴリとモダリティの結びつきを得ることができる．

提案手法は，確率モデル LDA の拡張であり，文献 [中村 08] と同様に確率的に様々な推論が可能となる．例えば，ロボットは物体を見ることで，その視覚情報から，その物体の聴覚情報や触覚情報の予測が可能となる．また，対話により概念と単語が結びつくため，ロボットが見た物を単語で表現することが可能となる．さらに，モダリティと単語の結びつきが獲得されるため，単語から特定のモダリティへ注意を向けること等も可能となる．

関連研究として，視覚情報のみを用いた物体カテゴリの教師なし学習に関する研究がある [Sivic 05, Fergus 03, Fei-Fei 05, Wang 09]．しかしカテゴリは，常に視覚的な情報のみで決定できるわけではなく，人間が物体をカテゴリに分類する際は，

連絡先: 中村 友昭，電気通信大学電子工学専攻，〒 182-8585
東京都調布市調布ヶ丘 1-5-1，naka_t@apple.ee.uec.ac.jp

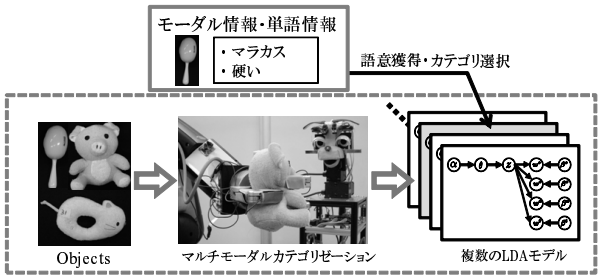

図 1: 提案手法の概要

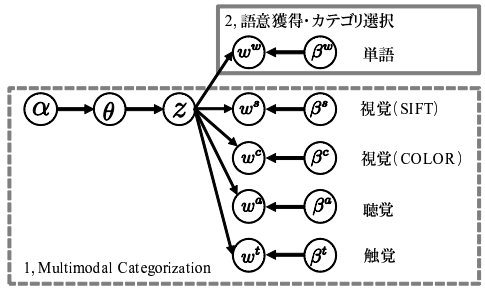

図 2: グラフィカルモデル

他にも様々な情報を用いていると思われる．また，本研究は言語獲得に関する研究 [Roy 02, Iwahashi 07, Yu 04] とも関連があると考えられる．しかし，これらの研究では触覚情報は用いておらず，モダリティ間の予測などは考えられていない．また，確率的にカテゴリ数を選択する研究も行われている [Attias 99, Cordueanu 01, Blei 06]．これらの研究では特定のカテゴリ数を一意に決めるもので，様々なカテゴリ数を選択する本研究とは異なる．また，本稿のようにモダリティとの結びつきを求めることは行われていない．

本稿の構成は以下の通りである．まず，2. で物体概念の形成を行うための，特定のモダリティとの結びつきを考慮した重み付きマルチモーダルカテゴリゼーションについて述べ，3. では，2. で形成した様々な概念に対して語意を接地し，その単語が表すカテゴリを選択する手法を述べる．次に実際にロボットを用いて概念の形成とカテゴリ選択に関する実験を 4. で述べ，最後に 5. で本論文をまとめる．

2. 概念学習

2.1 概念のグラフィカルモデル

ロボットは実際に物体を観察して得られるマルチモーダル情報をカテゴリ分類することで，様々な概念の形成を行う．図 2 の破線矩形内がマルチモーダルカテゴリゼーションのグラフィカルモデルとなる．まずこの破線矩形内を学習することで，マルチモーダル情報のカテゴリ分類を行う．w^s，w^c, w^a，w^t はそれぞれ視覚情報 (SIFT)，視覚情報 (色)，聴覚情報，触覚情報を示している．各情報の詳細については後で述べる．また z は物体のカテゴリを表している．さらに，w^s，w^c，w^a，w^t は，それぞれ多項分布 β^s，β^c，β^a，β^t から発生する．また，カテゴリ z の出現確率分布を表す多項分布のパラメータを θ とする．このパラメータは，ハイパーパラメータ α により決まるディリクレ事前分布に従う．

2.2 マルチモーダル情報

ここでは，ロボットが取得するマルチモーダル情報（視覚・聴覚・触覚情報）について述べる．

視覚情報

ロボット (図 3 左) は頭部にカメラを搭載しており，物体を観察することで得られる画像を視覚情報として利用する．画像は，物体毎に複数枚取得する (後に示す実験では，各物体に対して 50 枚の画像を用いた)．各画像から抽出する特徴量として，36 次元の PCA SIFT[Ke 04] を用いる．これは局所的な特徴である SIFT[Lowe 04] における記述子の代わりに，特徴点周辺の画素値に対して PCA を行い，上位 36 個の主成分を 36 次元の記述子として用いる．PCA SIFT は，SIFT に比べ高い表現力があることが知られている．

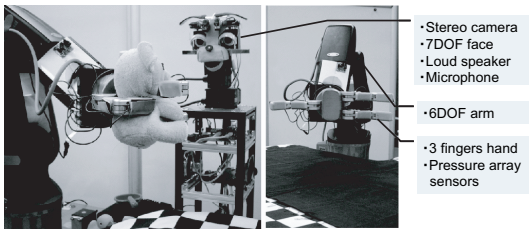

図 3: ロボット

PCA SIFT により，1 つの画像から多数の特徴ベクトルを得ることができる．特徴ベクトルの数は，画像により異なるため，このままでは物体の特徴量としては扱いにくい．そこで，これらの特徴ベクトルは，500 の代表ベクトルによりベクトル量子化することで，500 次元のヒストグラムとした．

さらに，2 つ目の視覚情報として，HSV 表色系の色相と彩度のヒストグラムを用いた．色相，彩度のビンの数はそれぞれ 10 とし，合計 100 次元のヒストグラムとした．

聴覚情報

聴覚情報として，物体を振ることで発生した音をマイクにより取得する．ひとつの物体を観測している間に得られる音声信号をフレームに分割し，フレーム毎の特徴量に変換する．特徴量として，13 次元の MFCC(Mel-Frequency Cepstrum Coefficient) を用いた．これにより，各フレームは 13 次元の特徴ベクトルとなる．最終的にこの特徴ベクトルも，ベクトル量子化を行い，50 次元のヒストグラムとなる．

触覚情報

触覚情報の取得には，3 本指のロボットハンドに取り付けられた 162 個のセンサーから構成されている触角アレイセンサーを用いた (図 3 右)．ロボットが実際に物体を把持することで，得られるセンサーの情報から触覚情報を計算する．ロボットが物体を把持することで得られるセンサーの時系列データの近似を行い，その近似パラメータを各センサーの特徴ベクトルとして扱う [中村 10]．さらに，この特徴ベクトルをベクトル量子化することで，15 次元のヒストグラムを触覚情報として用いる．

2.3 重み付きマルチモーダルカテゴリゼーション

ここでのカテゴリゼーションの問題は，図 2 のグラフィカルモデルの破線矩形内のパラメータ α，β^* を，ロボットが取得したマルチモーダル情報を用いて学習することに相当する．モデルのパラメータの学習は，与えられた学習データに対する目的関数を最大とするパラメータの推定によって実現される．

前述のように，カテゴリは様々なモダリティとの結びつきを考慮し，様々な粒度で形成される必要がある．そこで，カテゴリ数を N とし，さらに各モダリティへの重みとして，λ^s，λ^c，λ^a，λ^t を導入し，以下の対数尤度を考える．ただし，これらのパラメータの集合を $\boldsymbol{m} = \{N, \lambda^s, \lambda^c, \lambda^a, \lambda^t\}$ とする．

$$\begin{aligned}\mathcal{L}(\boldsymbol{w}^s, \boldsymbol{w}^c, \boldsymbol{w}^a, \boldsymbol{w}^t | \alpha, \beta^s, \beta^c, \beta^a, \beta^t, \boldsymbol{m}) \\ = \lambda^s \log P(\boldsymbol{w}^s | \alpha, \beta^s, N) + \lambda^c \log P(\boldsymbol{w}^c | \alpha, \beta^c, N) \\ + \lambda^a \log P(\boldsymbol{w}^a | \alpha, \beta^a, N) + \lambda^t \log P(\boldsymbol{w}^t | \alpha, \beta^t, N) \quad (1)\end{aligned}$$

あるモデルパラメータ $\boldsymbol{m}$ の下で，この対数尤度が最大となる α，β^* を推定することで，物体のカテゴリを形成することがで

きる．また，その分類はモデルのパラメータ $\boldsymbol{m}$ によって変化する．モダリティへの重み λ^* は，特定のモダリティとの結びつきの強さをあらわしており，この値により特定のモダリティと結びついたカテゴリを形成することができる．また，その分類の粒度は N によって変化する．しかし，この対数尤度は直接計算することが困難であるため，変分ベイズ法を適用する．変分ベイズ法は，対数尤度の計算が困難である場合，Jensen の不等式を適用することにより，対数尤度の下限値を最大化する手法である．あるモーダル情報を $\boldsymbol{w}^*$ とすると，$P(\boldsymbol{w}^*|\alpha,\beta^*,N)$ の下限値は次のようになる．

$$\begin{aligned}
&\log P(\boldsymbol{w}^*|\alpha,\beta^*,N)\\
&= \log\int\sum_{\boldsymbol{z}}\frac{P(\theta,\boldsymbol{z},\boldsymbol{w}^*|\alpha,\beta^*,N)}{Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N)}Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N)d\theta\\
&\geq \int\sum_{\boldsymbol{z}}Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N)\log P(\theta,\boldsymbol{z},\boldsymbol{w}^*|\alpha,\beta^*,N)d\theta\\
&\quad -\int\sum_{\boldsymbol{z}}Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N)\log Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N) \quad (2)
\end{aligned}$$

ここで $Q(\theta,\boldsymbol{z}|\gamma,\phi^*,N)$ は，$P(\theta,\boldsymbol{z}|\boldsymbol{w}^*,\alpha,\beta^*,N)$ を近似するために導入された確率分布であり，互いに独立な項の積で表されると仮定している．但し，ϕ^* は，それぞれのモーダル情報からカテゴリが発生する確率を表す多項分布のパラメータ，γ は θ のディリクレ事前分布のパラメータである．この式を式 (1) に代入し，式 (1) を最大化する α と β^* を EM アルゴリズムにより求めることとなる．EM アルゴリズムの更新式を以下に示す．

EM アルゴリズム

ランダムな初期値から初め，以下の E ステップと M ステップを収束するまで繰り返す．

[E ステップ]

各物体 d に関して以下の $\phi^*_{dw^*k}$ と γ_k を以下の式に従い更新する．

$$\phi^*_{dw^*k} \propto \beta^*_{kw^*}\exp\left(\psi(\gamma_{dk})-\psi(\sum_{k'}\gamma_{dk'})\right) \quad (3)$$

$$\begin{aligned}
\gamma_{dk} &= \alpha_k + \bar{\lambda}^s\sum_{w^s}\phi^s_{dw^sk} + \bar{\lambda}^c\sum_{w^c}\phi^c_{dw^ck}\\
&\quad + \bar{\lambda}^a\sum_{w^a}\phi^a_{dw^ak} + \bar{\lambda}^t\sum_{w^t}\phi^t_{dw^tk} \quad (4)
\end{aligned}$$

ただし，$\bar{\lambda^*} = \frac{\lambda^*}{\lambda^s+\lambda^c+\lambda^a+\lambda^t}$ とする．

[M ステップ]

$$\beta^*_{kw^*} \propto \sum_d n_{dw^*}\phi^v_{dw^*k} \quad (5)$$

$$\begin{aligned}
\frac{\partial L}{\partial\alpha_k} &= N\left(\psi(\sum_{k'}\alpha_{k'})-\psi(\alpha_k)\right)\\
&\quad + \sum_d\left(\psi(\gamma_{dk})-\psi(\sum_{k'}\gamma_{dk'})\right)=0 \quad (6)
\end{aligned}$$

ただし，n_{dw^*} は物体 d における特徴量 w^* の生起回数を，$d=1,...,N$ は物体のインデックスを，k はカテゴリのインデックスを表している．また，α_k は，式 (6) の L が最大となるパラメータを選択する．L の最大化には, Newton-Raphson 法を用いた反復計算を行う．

ここでは，パラメータ $\boldsymbol{m}$ を変化させ，様々なカテゴリを学習する．すなわち，複数のモダリティと結びついた物体概念や，特定のモダリティと結びついた抽象的な概念をあらわすモデルが学習される．これは，図 1 の破線矩形内に対応し，特定のモダリティと結びついた，様々な粒度のモデルが学習される．

表 1: 語意の学習に用いた単語

ぬいぐるみ	ガラガラ	コップ	ゴム人形
スポンジ	ペットボトル	ボール	マラカス
飲み物	黄色	楽器	丸い
硬い	柔らかい	積み木	茶色
緑色			

3. 語意の獲得とモデル選択

これまで述べた提案手法により，ロボットは教師なしで様々な概念を形成することが可能となった．ここでは，形成された概念に対して，単語の意味を接地し，さらに単語が表すカテゴリの選択を行う (図 1 実線矩形内)．提案するモデルは，図 2 であり，マルチモーダルカテゴリゼーションをさらに拡張したモデルとなっている．語意の獲得は，モデルの実線矩形内のパラメータを推定することに相当する．

3.1 語意の学習

図 2 の破線部矩形内は，前述のマルチモーダルカテゴリゼーションにより学習済みとなる．このモデルでは w^w は単語情報を表し，Bag-of-words モデルとして扱う．つまり，知覚情報と同様に，単語は発生位置に関係なく，その発生頻度でモデル化され，多項分布 β^w から発生する．人がロボットに物体を見せながら，その物体が含まれるような概念 (物体カテゴリ・色・形・硬さ等) を表す単語を教えることで語意の学習を行う．学習済みのマルチモーダルカテゴリゼーションのモデルを使用することで，ロボットは物体を見ることで，そのカテゴリ z を推定することができる．推定した結果と，その単語の生起回数から単語の発生確率を表す多項分布 β^w を，各モデル $\boldsymbol{m}$ に対して以下のように直接計算する．

$$P(w|z,\boldsymbol{m}) \propto \sum_i P(z|\bar{\boldsymbol{w}}^s_i,\boldsymbol{m})n(w,\bar{\boldsymbol{w}}^s_i) \quad (7)$$

ただし，$\bar{\boldsymbol{w}}^s_i$ は，語意の学習に用いた i 番目の物体の視覚特徴を表し，$n(w,\bar{\boldsymbol{w}}^s_i)$ は物体の視覚特徴 $\bar{\boldsymbol{w}}^s_i$ に対する単語 w の発生回数である．$P(z|\bar{\boldsymbol{w}}^s_i)$ は学習済みのマルチモーダルカテゴリゼーションによりカテゴリ認識をした結果である．なお上記の例では，視覚情報から単語の予測を行っているが，他のモダリティからの予測を用いて学習することも可能である．

3.2 カテゴリ選択

前章のマルチモーダルカテゴリゼーションでは，モデルパラメータ $\boldsymbol{m}$ を様々に変化させることで，様々な概念を構築した．ここでは，語意の獲得と同時に，単語が表すカテゴリとモデルパラメータの選択を行う．単語が特定のカテゴリからのみ多く発生していれば，その単語はそのカテゴリを正しく表現しているといえる．すなわち，あるモデル $\boldsymbol{m}$ から，カテゴリ z と単語 w が共起する確率 $P(w,z|\boldsymbol{m})$ が高いほど，単語 w がカテゴリ z をあらわしていると考えられ，以下のように単語 w が含まれるモデル $\boldsymbol{m}_w$ と単語 w が表すカテゴリ z_w を計算する．

$$(z_w,\boldsymbol{m}_w) = \underset{z,\boldsymbol{m}}{\operatorname{argmax}}\,P(w,z|\boldsymbol{m}) \quad (8)$$

4. 実験

実験では，33 個の物体を使用し，正しく人間の感覚に即したカテゴリを獲得することができるか検証した．

まず、カテゴリ数を 2 から 10 まで変化させ，さらにモダリティとの結びつき λ^* を様々に変化させ合計 136 個のモデルを学習した．次に，33 個の物体からランダムに 20 個の物体を選択し，ロボットに各物体の視覚情報を与えながら，その物体の特徴を表す単語を与えた．ロボットは，視覚情報のみからその

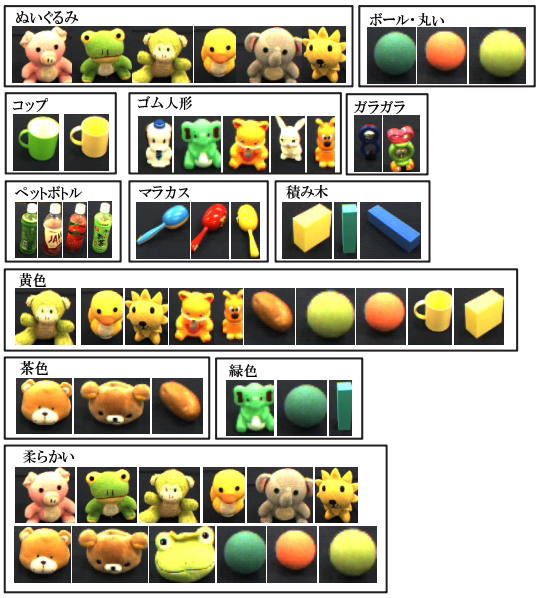

図 4: カテゴリ学習の結果

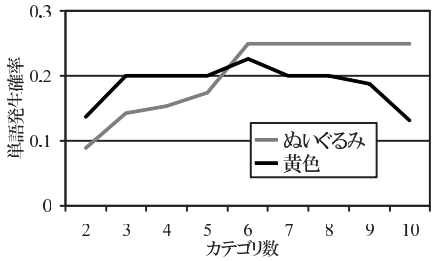

図 5: カテゴリ数と単語発生確率

カテゴリの予測を行い，単語の学習を行った．学習に用いた単語は表 1 の通りである．

獲得された概念の一部を図 4 に示す．ぬいぐるみやマラカスといった物体概念を表す概念が正しく獲得されていることが分かる．また，学習時には視覚情報のみしか与えていないにもかかわらず，色概念や触覚概念も概ね正しく獲得されている．図 5 に，カテゴリ数 N の時，単語 w が発生する確率を示した．ただし，カテゴリ数 N の時の単語 w の発生確率は以下のように計算する．

$$P_{max}(w|N) = \max_{z,\lambda^s,\lambda^c,\lambda^a,\lambda^t} P(w, z|\boldsymbol{m}) \tag{9}$$

単語「ぬいぐるみ」の発生確率は，カテゴリ数が増加するにつれて高まっていることが分かる．これは，カテゴリ数が小さいと，ぬいぐるみ以外の物体が入り込んでくるため，単語の発生確率が減少しているためである．また,「ぬいぐるみ」と比較してより抽象的な概念である「黄色」では，カテゴリ数が 5 で最大となり，それよりも少なくても，多くても単語「黄色」の発生確率が減少している．これは，カテゴリ数が 5 未満であると，黄色い物体以外が入り込んでしまい，5 より大きくなると黄色い物体が複数のカテゴリに分かれてしまうためである．このように，人と各物体に関する対話を行い，語意を獲得する過程でその単語が表すカテゴリを特定することができる．

5. まとめ

本稿では，複数のマルチモーダル LDA を用いて様々な概念を形成し，人からの教示で，人の感覚に即したカテゴリの選択を行った．提案手法では，全ての物体の全てのモーダル情報が与えられなくとも，一部の情報のみから予測を用いて，他のモダリティの概念を表す単語を獲得することが可能となった．このよう一部の情報からでも予測を用いて学習できる枠組みは，ロボットが人の概念を獲得するうえで重要であると考えられる．実験の結果から概ね正しいカテゴリが形成されたことが分かる．今回の実験では，定量的な評価を行うことができなかったが，今後正解となるカテゴリを定め定量的な評価を行うこと考えている．また，今回単語の教示は 1 人で行ったが，複数人で単語を教示した場合にどのようなカテゴリが形成されるかを検証する必要があると考えている．

謝辞

本研究は，科研費（20500186，20500179）及び新学術領域研究「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

[Sivic 05] Sivic,J., Russell,B.C., Efros,A.A., Zisserman,A., and Freeman,W.T.: “Discovering Object Categories in Image Collections”, *AI Memo*, 2005-005, pp.1-12, (2005)

[Fergus 03] Fergus,R., Perona,P., and Zisserman,A.: “Object Class Recognition by Unsupervised Scale-invariant Learning”, *Proc. of CVPR2003*, Vol.2, pp.264-271 (2003)

[Fei-Fei 05] Fei-Fei,L., and Perona,P.: “A Bayesian Hierarchical Model for Learning Natural Scene Categories”, *Proc. of CVPR2005*, Vol.2, pp.524- 531, (2005)

[Wang 09] Wang,C., Blei,D., and Fei-Fei,L.: “Simultaneous Image Classification and Annotation”, *Proc. of CVPR2009*, (2009)

[中村 08] 中村 友昭，長井 隆行，岩橋 直人: “ロボットによる物体のマルチモーダルカテゴリゼーション”, 電子情報通信学会論文誌 D，Vol.J91-D No.10，pp.2507-2518 (2008)

[Hofman 01] Hofmann,T.: “Unsupervised Learning by Probabilistic Latent Semantic Analysis”, *Machine Learning*, Vol.42, pp.177-196 (2001)

[Blei 03] Blei,D.M, Ng,A.Y, and Jordan,M.I.: “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, pp.993-1022 (2003)

[Roy 02] Roy,D., and Pentland,A.: “Learning Words from Sights and Sounds: A Computational Model”, *Cognitive Science*, Vol.26, No.1, pp.113-146 (2002)

[Iwahashi 07] N.Iwahashi: “Robots That Learn Language: A Developmental Approach to Situated Human-Robot Conversations”, *In N.Sankar ed. Human-Robot Interaction*, pp.95-118, I-Tech Education and Publishing (2007)

[Yu 04] Yu,C., and Ballard,D.: “On the Integration of Grounding Language and Learning Objects”, *Proc. of 19th National Conference on Artificial Intelligence(AAAI)*, pp.488-494 (2004)

[Attias 99] Attias,H.: Inferring Parameters and Structure of Latent Variable Models by Variational Bayes, *Proc. of 15th Conference on Uncertainty in Artificial Intelligence*, pp.21-30 (1999)

[Cordueanu 01] Corduneanu,A., and Bishop,C,M.: Variational Bayesian Model Selection for Mixture Distributions, *Proc. of International Conference on Artificial Intelligence and Statistics* , pp.27-34 (2001)

[Blei 06] Blei,D.M. and Jordany,M.I.: Variational Inference for Dirichlet Process Mixtures *Journal of Bayesian Analysis*, Vol.1(1), pp.121-144 (2006)

[Ke 04] Ke,Y., and Sukthankar,R.: “PCA-SIFT: A More Distinctive Representation for Local Image Descriptors”, *Proc. of Computer Vision and Pattern Recognition* (2004)

[Lowe 04] Lowe,D.G.: “Distinctive image features from scale-invariant keypoints”, *Int. Journal of Computer Vision*, 60(2), pp. 91-110 (2004)

[中村 10] 中村 友昭，西田 匡志，長井 隆行: “把持動作による物体カテゴリの形成と認識”, 情報処理学会全国大会, 5V-3, (2010)