

# 注意を介する知覚体制化による構成的物体認識 -物体カテゴリの確率的学習モデルの構築-

A Preliminary Study on Attention-mediated Perceptual Organization and Learning for Constructive Object Recognition

渥美雅保

Masayasu Atsumi

創価大学工学部情報システム工学科

Dept. of Information Systems Sci., Faculty of Eng., Soka University

This paper proposes a probabilistic model of attention-mediated perceptual organization and learning for constructive object recognition. This model consists of the following three submodels: (1) a model of attention-mediated perceptual organization of adjacent segments which are segmented on dynamically-formed Markov random fields, (2) a model of local feature representation of segments by bags of key features, and (3) a model of semi-supervised learning of categorical objects and their context based on a probabilistic latent component analysis of cooccurrent segments within categories and their cross analysis among categories. This model makes it possible to learn categories of objects under incomplete instructions that those objects are exists somewhere in given crowded scenes.

## 1. はじめに

多くの物体が込入って置かれた情景において、人は情景内にある物体のカテゴリを識別することができる。しかし、人はそれら物体のカテゴリを必ずしも教師付きで一つ一つ学習してきたわけではない。本論文では、(a) 物体のカテゴリ名は半教師付き学習により獲得される、(b) 物体はそのコンテキストの中で統計的な分析を通じて連想学習される、という仮定のもとに、注意、セグメンテーション、知覚体制化の初期視覚過程と関連付けた半教師付き統計的物体学習のモデルを提案する。本モデルは、次の3つのサブモデルから構成される。第1は、注意を介するセグメントの知覚体制化 [Atsumi 09] のモデルである。第2は、BoF (Bag of Feature) [Csurka 04] によるセグメントの局所特徴量表現のモデルである。第3は、PLCA (Probabilistic Latent Component Analysis) [Hofmann 01, Shashanka 08] に基づく物体とそのコンテキストの半教師付き統計的学習のモデルである。第1の注意を介するセグメントの知覚体制化では、顕著性の高い点の周りに動的に形成されるマルコフ確率場で図地セグメンテーションが行われ、注意度の高いセグメントの近傍での知覚体制化により注意共起セグメントグラフが生成される。第2のセグメント特徴量の表現では、顕著点での SIFT 特徴量 [Lowe 04] を K-tree [Shlomo 00] によりクラスタリングしてえられたキー特徴量の集合を用いて、各セグメントの BoF がセグメントの局所特徴量として計算される。第3の物体とそのコンテキストの半教師付き統計的学習では、カテゴリごとに集められた込み入った情景内にそのカテゴリの物体が存在する画像集合の族から、各カテゴリの物体と関連コンテキストのクラスファイアが PLCA とカテゴリ内・カテゴリ間でのクラス特徴の確率分布の分析に基づき学習される。本学習方式により、与えられた情景内のどこかに指定されたカテゴリの物体が存在するという限定された半教師信号のもとで、当該の物体を統計的に学習することが可能となる。

以下、2章で注意を介するセグメントの知覚体制化と共起セグメントの BoF 特徴表現、3章で PLCA に基づく物体カテゴリの確率的学習について述べる。

## 2. 注意を介する知覚体制化

### 2.1 注意セグメントの体制化

注意を介する知覚体制化のモデル (図 1) は、画像の特徴空間上で計算される前注意のための顕著性マップ [Itti 98, Atsumi 07]、図地セグメンテーションのための動的に形成されるマルコフ確率場の集まり [Atsumi 09]、図セグメントの集合を保持し近接セグメント間で注意を介する知覚体制化により共起セグメントグラフを計算する視覚作業記憶、及びセグメントの選択的注意の機構から構成される。画像の特徴量としては、多重解像度ピラミッド上で明度・彩度・色相、それらのコントラストと方位成分が求められる。顕著性マップは、多重解像度ピラミッド上で明度コントラスト、彩度コントラスト、色相コントラスト、及び方位成分から多レベルで顕著性を計算し、それらを一つの顕著性マップに統合することで求められる。

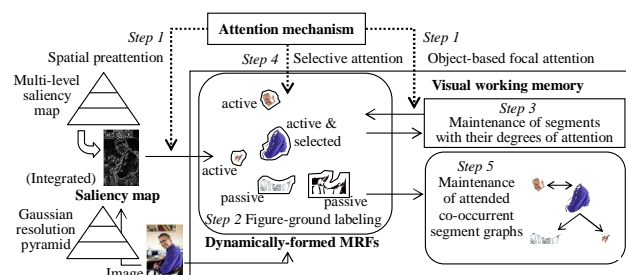


図 1: 注意を介する知覚体制化

注意を介する知覚体制化の手続きの概要を次に示す。

- Step 1 顕著性マップまたは視覚作業記憶から注意度に基づき確率的に前注意点またはセグメントを選択する。
- Step 2 前注意点またはセグメントの周囲に一定のマージンで明度と色相のマルコフ確率場を展開し図地ラベリングを行うことをある一定回数、またはセグメントが収束するまで繰り返す。ここで、図地ラベリングは平均場近似を用いた EM 手続き [Zhang 92] により遂行される。複数のセグメントが合併条件を満たす場合は、それらセグメントを合併する。

**Step 3** セグメントの注意度を, セグメントの顕著性, 閉領域性, 及び注意バイアスに基づき計算する [Atsumi 09]. ここで, 注意バイアスとは, 顔らしい領域への注意性向のような制約をいう. Step 1 で選択された前注意点またはセグメントを Step 2 で合併を伴って展開したセグメントの集まりを能動的記憶セグメント集合とする.

**Step 4** 能動的記憶セグメント集合から注意度の大きい既定数のセグメントを選択的注意セグメントとして選択する.

**Step 5** 視覚作業記憶において, 選択的注意セグメントとそれに隣接する能動的記憶セグメントの組を記憶し, それらの中で共起率の高い組を連結することにより, 選択的注意セグメントを幹ノードとし能動的記憶セグメントを葉ノードとする共起セグメントグラフの集合を求める.

知覚体制化により, 1つの物体を構成する複数のセグメントの体制化やある物体セグメントとそのコンテキストとの体制化がなされる.

## 2.2 共起セグメントの局所特徴の BoF 表現

共起セグメント, 及びその各セグメントの特徴を顕著点の局所特徴の BoF ヒストグラムにより表現する. そのために, まず, 共起セグメントの各セグメントから顕著性が大きい点を顕著点として選択し, その点の局所特徴として SIFT 特徴量を計算する. ここで, セグメントの顕著点としては, 多レベル顕著性マップの各レベルからセグメントの内部の点で顕著性がある閾値以上の点を選択する. また, SIFT 特徴量は, 回転に不変な 128 次元の正規化済み明度勾配特徴量で, その顕著点のレベルの解像度のもとで計算される. 次に, 全てのセグメントの顕著点の SIFT 特徴量の集合をクラスターリングしてキー特徴量の集合を求める. ここで, クラスターリングには K-tree を用いている. そして, 最後に, このキー特徴量の集合をコードブックとして用いて, セグメントの BoF, 及びそれらから共起セグメントの BoF を計算し, セグメント, 及び共起セグメントの特徴表現とする.

## 3. 物体カテゴリの確率的学習

与えられた情景内に指定されたカテゴリの物体が存在するという半教師信号のもとでそれらカテゴリの物体を学習するために, カテゴリごとに集められた込み入った情景内にそのカテゴリの物体が存在する画像集合の族から, 各カテゴリの物体, 及び同時にそれらカテゴリの物体と同時に出現しやすい関連コンテキストを統計的に学習する問題を考える. 具体的には, 複数のカテゴリの画像から注意を介する知覚体制化により分節された共起セグメントの集合に対して, まず, カテゴリごとにセグメントの BoF ヒストグラム集合からカテゴリ内のマルチクラスクラシファイアを PLCA に基づく方法により求め, 次に, カテゴリ間のクラスのクロス分析に基づきカテゴリの物体クラスとそのコンテキストクラスを求める. 以下, カテゴリ  $c$  の画像  $i$  から分節されたセグメント  $j$  を  $s_{c,i,j}$ , カテゴリ  $c$  の全画像から分節されたセグメントの集合を  $S_c$ , その数を  $N_{c_s}$ , キー特徴量集合  $F$  の  $n$  番目の要素を  $f_n$ , キー特徴量の数を  $N_f$ , セグメント  $s_{c,i,j}$  の BoF ヒストグラムを  $\{H(s_{c,i,j}, f_n)\}_{f_n \in F}$ , カテゴリ  $c$  の潜在クラスを  $q_{c,r}$ , その数を  $N_{c_q}$ , 潜在クラスの集合を  $Q_c$ , カテゴリの総数を  $N_c$  とする.

### 3.1 PLCA に基づくマルチクラスクラシファイア

カテゴリ  $c$  のセグメントのマルチクラスクラシファイアの学習は, カテゴリ  $c$  のすべてのセグメント  $s_{c,i,j}$  の BoF ヒス

トグラム  $\{H(s_{c,i,j}, f_n)\}_{f_n \in F}$  の集合に対して, 対数尤度

$$L_c = \sum_{i_j} \sum_n H(s_{c,i,j}, f_n) \log(p(s_{c,i,j}, f_n)) \quad (1)$$

を最大にする  $p(s_{c,i,j}, f_n) = \sum_r p(q_{c,r}) p(s_{c,i,j} | q_{c,r}) p(f_n | q_{c,r})$ , 即ち  $\{p(q_{c,r}) | q_{c,r} \in Q_c\}$ ,  $\{p(s_{c,i,j} | q_{c,r}) | s_{c,i,j} \in S_c, q_{c,r} \in Q_c\}$ ,  $\{p(f_n | q_{c,r}) | f_n \in F, q_{c,r} \in Q_c\}$ , 及び潜在クラスの数  $N_{c_q}$  を推定することによりなされる.

与えられた潜在クラスの数のもとでのこれら確率は EM アルゴリズムにより,

[E-step]

$$p(q_{c,r} | s_{c,i,j}, f_n) = \frac{[p(q_{c,r}) p(s_{c,i,j} | q_{c,r}) p(f_n | q_{c,r})]^\beta}{\sum_{q_{c,r'}} [p(q_{c,r'}) p(s_{c,i,j} | q_{c,r'}) p(f_n | q_{c,r'})]^\beta} \quad (2)$$

[M-step]

$$p(f_n | q_{c,r}) = \frac{\sum_{s_{c,i,j}} H(s_{c,i,j}, f_n) p(q_{c,r} | s_{c,i,j}, f_n)}{\sum_{f_{n'}} \sum_{s_{c,i,j}} H(s_{c,i,j}, f_{n'}) p(q_{c,r} | s_{c,i,j}, f_{n'})} \quad (3)$$

$$p(s_{c,i,j} | q_{c,r}) = \frac{\sum_{f_n} H(s_{c,i,j}, f_n) p(q_{c,r} | s_{c,i,j}, f_n)}{\sum_{s_{c,i,j'}} \sum_{f_n} H(s_{c,i,j'}, f_n) p(q_{c,r} | s_{c,i,j'}, f_n)} \quad (4)$$

$$p(q_{c,r}) = \frac{\sum_{s_{c,i,j}} \sum_{f_n} H(s_{c,i,j}, f_n) p(q_{c,r} | s_{c,i,j}, f_n)}{\sum_{s_{c,i,j}} \sum_{f_n} H(s_{c,i,j}, f_n)} \quad (5)$$

を繰り返すことにより推定される. ここで,  $\beta$  は温度係数である.

潜在クラスの数の決定は, 1つまたはある少数の潜在クラスから初めて, 一定回数を上限とする EM の繰返し毎に各クラス内の散らばり度

$$\delta_{c,r} = \frac{\sum_{s_{c,i,j}} (\sum_{f_n} |p(f_n | q_{c,r}) - D(s_{c,i,j}, f_n)|) \times p(s_{c,i,j} | q_{c,r})}{N_{c_s}} \quad (6)$$

ここで,

$$D(s_{c,i,j}, f_n) = \frac{H(s_{c,i,j}, f_n)}{\sum_{f_{n'}} H(s_{c,i,j}, f_{n'})} \quad (7)$$

を求め, それら値がある閾値以上の潜在クラスのうち最大値をとるクラスを 2 分割して EM 手続きを続けることを, すべてのクラスの  $\delta_{c,r}$  が閾値以下になるまで繰り返すことによりなされる. 潜在クラスの分割は, 元のクラスを  $q_{c,r}$ , 分割後のクラスを  $q_{c,r_1}$ ,  $q_{c,r_2}$  とするとき, 2つのセグメントを選択確率  $\{p(s_{c,i,j} | q_{c,r})\}$  のもとで確率的に取り出し, それら 2つのセグメント  $s_{c,1}$ ,  $s_{c,2}$  の BoF ヒストグラム  $H(s_{c,1}, f_n) = [h_{c,1}(1), \dots, h_{c,1}(N_f)]$ ,  $H(s_{c,2}, f_n) = [h_{c,2}(1), \dots, h_{c,2}(N_f)]$  を用いて, クラス  $q_{c,r}$  を

$$p(f_n | q_{c,k}) = \frac{h_{c,k}(n) + \alpha}{\sum_{n'} h_{c,k}(n') + \alpha} \quad (8)$$

$$p(s_{c,i,j} | q_{c,k}) = p(s_{c,i,j} | q_{c,r}) \quad (9)$$

$$p(q_{c,k}) = \frac{p(q_{c,r})}{2} \quad (10)$$

( $k = r_1, r_2$ ) のクラス  $q_{c,r_1}$ ,  $q_{c,r_2}$  に置き換える. ここで,  $\alpha$  は正値の補正係数である. すべての潜在クラスの散らばり度  $\delta_{c,r}$  が閾値以下になりクラス数が決定するまでは温度  $\beta$  を 1.0 に固定し, その後, 1.0 から与えられたスケジュールの元で段階的に減らして EM 手続きを収束するまで実行する.

### 3.2 物体カテゴリの学習

各カテゴリの物体クラスとコンテキストクラスの学習では、当該カテゴリに高い頻度で出現し、かつその他のカテゴリには低い頻度でしか出現しない潜在クラスを物体を構成するセグメントのクラスとして選択し、それらから物体クラスのキー特徴量の条件付き確率分布を合成する。また、これにより同時にそのカテゴリで物体と共起しやすいコンテキストクラスが選別される。そのために、まず、カテゴリ間でクラスの同値関係、即ち同一視可能なクラスの対応関係を分析する。次に、各カテゴリの各クラスについて、それが物体を構成するセグメントのクラスであるか否かを判別する。そして、物体を構成するクラスの集まりからそのカテゴリの物体クラスのキー特徴量の条件付き確率分布を求める。

カテゴリ間での潜在クラスの同値関係の分析では、異なるカテゴリのクラスのキー特徴量の条件付き確率分布間で確率分布距離を計算し、各カテゴリの各クラスに関して、それとの確率距離がある与えられた閾値以下のクラスを同一視可能なクラスとして枚挙する。ここで、任意のカテゴリ  $c_1$  のクラス  $r_1$  とカテゴリ  $c_2$  のクラス  $r_2$  の間のキー特徴量の条件付き確率分布距離は、

$$J(q_{c_1, r_1}, q_{c_2, r_2}) = \sum_{f_n} (p(f_n | q_{c_1, r_1}) - p(f_n | q_{c_2, r_2})) \log \left( \frac{p(f_n | q_{c_1, r_1})}{p(f_n | q_{c_2, r_2})} \right) \quad (11)$$

で与えられる。

カテゴリ  $c$  の潜在クラス  $q_{c,r}$  と同じと判定された異なるカテゴリ  $c_k$  のクラスの集合を  $Q(q_{c,r}) = \{q_{c_k, r_k}\}$ 、その数を  $|Q(q_{c,r})|$  とする。このとき、クラス  $q_{c,r}$  が物体を構成するクラス、即ち当該カテゴリに高い頻度で出現し、かつその他のカテゴリには低い頻度でしか出現しないクラスであることを表す指標として、カテゴリクラス度

$$\gamma(q_{c,r}) = (1 - \mu^{-(N_{c_q} \times p(q_{c,r}))}) - \frac{\sum_{q_{c_k, r_k} \in Q(q_{c,r})} p(q_{c_k, r_k})}{|Q(q_{c,r})|} \quad (12)$$

を定義する。ここで、 $\mu (> 1)$  は、カテゴリ内で当該クラスの出現が平均的なときの式 (12) の第 1 項の値を定める係数である。このとき、クラス  $q_{c,r}$  が物体を構成するクラスかコンテキストクラスかの判定は、そのクラスのカテゴリクラス度がある与えられた閾値以上か否かで行われる。

カテゴリ  $c$  の物体クラスのキー特徴量の条件付き確率分布は、物体を構成するクラスの集合を  $Q_c^* = \{q_{c,k}\}$ 、合成される物体クラスを  $q_{c,o}$  とするとき、

$$p(f_n | q_{c,o}) = \sum_{q_{c,k} \in Q_c^*} (\lambda(q_{c,k}) \times p(f_n | q_{c,k})) \quad (13)$$

$$\lambda(q_{c,k}) = \frac{p(q_{c,k})}{\sum_{q_{c,k'} \in Q_c^*} p(q_{c,k'})} \quad (14)$$

で求められる。

## 4. おわりに

本稿では、注意を介するセグメントの知覚体制化のもとで共起セグメントの BoF 特徴から PLCA に基づき物体のカテゴリを学習する一般物体認識のための確率モデルについて述べた。図 2 に画像データベース [Griffin 07] の画像を用いた予備的実験の一部を示す。本モデルの画像データベースの画像を用いた評価の詳細に関しては別稿でまとめることとしたい。

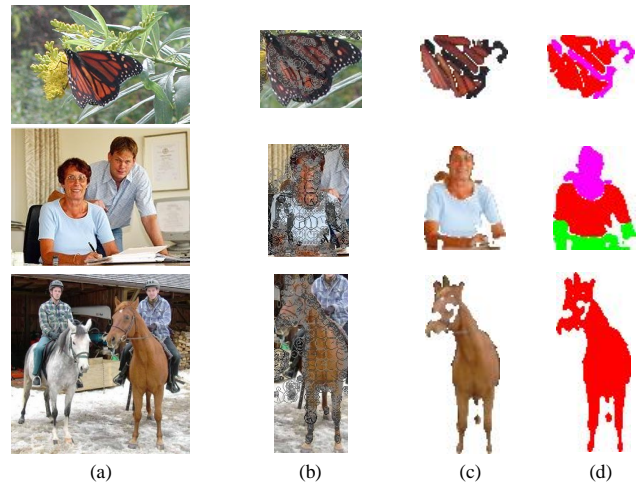


図 2: 予備的実験結果: (a) 画像, (b) 顕著点 SIFT 特徴記述, (c) カテゴリ物体の共起セグメント, (d) セグメントラベル (色分け)

## 参考文献

- [Atsumi 07] Atsumi, M.: Stochastic Attentional Selection and Shift on the Visual Attention Pyramid, Proc. of the 5th International Conference on Computer Vision Systems, CD-ROM, 10p. (2007)
- [Atsumi 09] Atsumi, M.: A Probabilistic Model of Visual Attention and Perceptual Organization for Constructive Object Recognition, Advances in Visual Computing (5th International Symposium, ISVC2009 Proceedings), (G. Bebis et al. (Eds.)), LNCS Vol.5876 Part II, pp.778-787, Springer-Verlag Berlin Heidelberg (2009)
- [Csurka 04] Csurka, G., Bray, C., Dance, C. and Fan, L.: Visual Categorization with Bags of Keypoints, Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp.1-22 (2004)
- [Griffin 07] Griffin, G., Holub, AD., Perona, P.: The Caltech-256, Caltech Technical Report (2007)
- [Hofmann 01] Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis, Machine Learning, Vol.42, pp.177-196 (2001)
- [Itti 98] Itti, L., Koch, C., Niebur, E.: A Model of Saliency-based Visual Attention for Rapid Scene Analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.20, pp.1254-1259 (1998)
- [Lowe 04] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, Vol.60, No.2, pp.91-110, (2004)
- [Shashanka 08] Shashanka, M., Raj, B., Smaragdis, P.: Probabilistic Latent Variable Models as Nonnegative Factorizations, Computational Intelligence and Neuroscience, Vol.2008, Article ID 947438, 8p. (2008)
- [Shlomo 00] Shlomo, G.: K-Tree; A Height Balanced Tree Structured Vector Quantizer, Proc. of the 2000 IEEE Signal Processing Society Workshop, Vol.1, pp.271-280 (2000)
- [Zhang 92] Zhang, J.: The Mean Field Theory in EM Procedures for Markov Random Fields, IEEE Trans. on Signal Processing, Vol.40, No.10, pp.2570-2583 (1992)