

注意と共起に基づくシーンの学習と認識

Scene Learning and Recognition based on Attention and Co-occurrence

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 describes a method of object categorization in context based on attention-guided co-occurrence and the probabilistic latent component tree with boosted features. In this method, object classes are obtained for each scene category, a classification tree with boosted features is generated for all the object classes and co-occurrence is analyzed among object categories in scenes. Through experiments using images of plural categories in image data sets, it is shown that the method analyzes intra-categorical composition and inter-categorical difference of scene categories and improves object categorization performance by using boosted features and co-occurrence.

1. はじめに

視覚的注意は多くの物体から構成される外界から選択的に物体を切り出すことにより視覚世界を形成し、視覚世界の物体に関する知識は物体のコンテキストと物体の見えの類似性により体系化されてカテゴリラベル付けされているとみなしうる。このとき、物体のコンテキストは物体カテゴリの共起関係により表現され、物体の見えの類似性は見え特徴の分類木により表現されうる。本論では、この観点から、注意と共起のもとの情景内物体の分類木に基づく認識の問題を扱う。このために、次の3つの学習問題を設定する。第1に、選択的注意が向けられたセグメント集合から物体と情景の特徴を学習する問題、第2に、物体の見え特徴の分類木を生成し頑強な識別のための特徴を選択する問題、第3に、情景を構成する物体間の共起性を学習する問題である。ここで、物体を頑強に識別するための特徴選択も注意の1つの働きによると考えられる。これら問題に対して、「注意にガイドされた共起とブースト特徴を埋め込んだ確率潜在コンポーネント木手法」を提案する。本手法では、情景内の複数の物体の1つにラベルがつけられた情景画像の集合が学習のために与えられる。ここで、情景内のラベル付き物体はその情景の前景とみなされる物体であり、その他の後景物体にはラベルがつけられていない。前景物体が同じラベルを持つ情景画像の集合は情景カテゴリを形成し、情景画像はその中のどの物体が前景とみなされるかに依存して複数の情景カテゴリに含まれる。

関連する研究として、顕著性に基づく注意 [Itti 01] のもとで物体を検出・認識する手法 [Walther 05, Atsumi 08, Liu 11]、物体や情景のカテゴリゼーションに確率潜在変数モデルを用いる手法 [Bosch 06, Huang 09] やカテゴリゼーションを階層化する手法 [Sivic 08]、物体のカテゴリゼーションに物体間の共起を用いる手法 [Galleguillos 08, Choi 12]、画像分類にブースティングを用いる手法 [Tu 05, Antenreiter 06] がある。提案手法の主な特徴は、空間前注意と物体焦点注意を動的に形成されるマルコフ場での図地セグメンテーションでリンクしている点、選択的に注意されたセグメント集合を共起の学習に結び付けることを可能としている点、分類木上で信頼度付き特徴選択が事後的にできる点、共起制約のもとで情景内の複数の物体が同時に認識され、また、前景物体の推測ができる点である。

2. 提案手法

情景カテゴリの集合を C とする。情景カテゴリ $c \in C$ の画像は、そのカテゴリの物体を前景に、他のカテゴリの物体を後景に含む画像である。情景カテゴリ c の画像 i から分節された j 番目の物体セグメントを $s_{c,i,j}$ 、情景カテゴリ c の任意の画像から分節された物体セグメントの集合を S_c とする。

2.1 注意にガイドされた情景の解析

注意にガイドされたセグメント集合の抽出 (図 1) は、画像の顕著性マップから選ばれた複数の高顕著度の前注意点の周りに動的に形成されるマルコフ確率場でのセグメンテーションと、それらセグメントに対して定義される注意度を用いたセグメントの選択とグルーピングにより、次の手順でなされる [Atsumi 10]。

Step 1 顕著性マップから顕著度に基づき確率的に複数の前注意点を選択する。

Step 2 前注意点の周囲にあるマージンで徐々に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。ここで、セグメントの顕著性は、セグメントとその周囲の平均特徴量

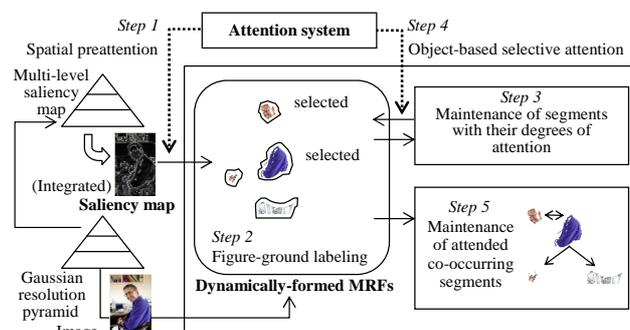


図 1: 注意にガイドされたセグメント集合の抽出

の差 (面顕著度) とセグメント内の各点の顕著度の最大値 (スポット顕著度) から求められる。閉領域性は、セグメントが画像内で閉じているかにより判定される。注意バイアスとは、顔らしい領域への注意性向のような先験または経験的な注意性向を与える制約である。

Step 4 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 選択的注意セグメントとそれに隣接するセグメントの集合を抽出する。

以上の手順を各情景カテゴリ $c \in C$ の画像に適用することにより、セグメントの集合 S_c を求める。

物体セグメントをその局所特徴の BoF (Bag of Features) により表現する。いま、コードブックとしてのキー特徴の集合を F 、 F の n 番目のキー特徴を f_n 、キー特徴の数を N_F とする。このとき、物体セグメント s_{c,i_j} はキー特徴の BoF により $H(s_{c,i_j}) = [h_{c,i_j}(f_1), \dots, h_{c,i_j}(f_{N_F})]$ と表される。

情景カテゴリ $c \in C$ に対して、それに含まれる画像の集合から求められた BoF の集合を $H_c = \{H(s_{c,i_j}) | s_{c,i_j} \in S_c\}$ とする。このとき、情景カテゴリ c の物体カテゴリを表すクラス集合 $Q_c = \{q_{c,r} | r = 1, \dots, N_{Q_c}\}$ を潜在クラス数可変の確率潜在コンポーネント解析により求める [Atsumi 10, Atsumi 12]。このために、BoF 集合 H_c に対して、対数尤度 $L_c = \sum_{i_j} \sum_n h_{c,i_j}(f_n) \log(p(s_{c,i_j}, f_n))$ 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、クラス確率 $\{p(q_{c,r}) | q_{c,r} \in Q_c\}$ 、条件付きインスタンス確率 $\{p(s_{c,i_j} | q_{c,r}) | s_{c,i_j} \in S_c, q_{c,r} \in Q_c\}$ 、条件付きキー特徴確率分布 $\{p(f_n | q_{c,r}) | f_n \in F, q_{c,r} \in Q_c\}$ 、及び潜在クラスの数 N_{Q_c} を推定する。ここで、キー特徴確率分布はクラスが表す物体カテゴリの見え特徴、クラス確率は情景カテゴリ内にある見えの物体カテゴリが存在する比率、インスタンス確率は物体セグメントがそのクラスのインスタンスである程度を与える。

これらの確率と潜在クラス数は、クラス分割を伴う Tempered EM アルゴリズムにより推定される。即ち、1つまたは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ばらつき度を求めて、ばらつき度が最大のクラスを2分割する。この過程を繰り返してクラス数を確定し、その後 Tempered EM により確率の値を求める。

情景カテゴリ c の特徴は、その物体カテゴリを表す各クラス $q_{c,r} \in Q_c$ のキー特徴確率分布をクラス確率の重み付けで合成したキー特徴確率分布 $\{p(f_n | Q_c) | f_n \in F\}$ により表される。

2.2 物体分類木の生成と特徴選択

全情景カテゴリから求められる物体クラスの全集合 $Q^* = \cup_{c \in C} Q_c$ に対して、確率潜在コンポーネント木をそれらクラスを葉に持つ分類木として生成する [Atsumi 12]。いま、あるブランチノードをルートとする部分木の葉ノードに割り当てられるクラスの集合を $Q^0 (\subseteq Q^*)$ とし、このブランチノードを $b(Q^0)$ と表す。このとき、まず、各キー特徴 $f_n \in F$ について、 Q^0 をキー特徴 f_n の確率値に応じて、2つの部分集合 $Q_{f_n}^1 = \{q_{c,r} | p(f_n | q_{c,r}) = 0, q_{c,r} \in Q^0\}$ と $Q_{f_n}^2 = \{q_{c,r} | p(f_n | q_{c,r}) > 0, q_{c,r} \in Q^0\}$ に分割する。次に、 $Q_{f_n}^1$ と $Q_{f_n}^2$ について、それぞれに含まれるクラスのキー特徴確率分布の平均分布 $\{\mu_{Q_{f_n}^1}(f_{n'}) | f_{n'} \in F\}$ と $\{\mu_{Q_{f_n}^2}(f_{n'}) | f_{n'} \in F\}$ を求める。そして、これら平均キー特徴確率分布に対して、次の KL 情報量に基づく距離指標

$$\delta_{f_n} = \sum_{q_{c,r} \in Q_{f_n}^1} p(q_{c,r}) \left(\sum_{f_{n'} \in F} p(f_{n'} | q_{c,r}) \log \frac{p(f_{n'} | q_{c,r})}{\mu_{Q_{f_n}^1}(f_{n'})} \right) \quad (1)$$

$$+ \sum_{q_{c,r} \in Q_{f_n}^2} p(q_{c,r}) \left(\sum_{f_{n'} \in F} p(f_{n'} | q_{c,r}) \log \frac{p(f_{n'} | q_{c,r})}{\mu_{Q_{f_n}^2}(f_{n'})} \right)$$

を計算する。最後に、すべてのキー特徴 $f_n \in F$ に対して、 δ_{f_n} の最小値を与えるクラス集合分割 Q^1, Q^2 を求める。以上の処理を、ルートノード $b(Q^*)$ から始めて、生成された子ブランチノードに関して葉ノードが生成されるまで再帰的に繰り返す。葉ノードには1つの物体クラス $q_{c,r}$ が割り当てられ、クラス確率、キー特徴確率分布、及びインスタンス確率が保持される。また、ブランチノードにも、その2つの子ノードから合成されるクラス確率とキー特徴確率分布が保持される。葉ノードの物体クラスのカテゴリラベルは、物体クラスのインスタンス確率最大の物体セグメントに順次与えられる物体カテゴリラベルを用いて半教師付きでラベル付けされる。

確率潜在コンポーネント木の各ブランチノードにおいて、与えられた物体セグメントのカテゴリが2分岐のどちらにあるかの判定に用いるキー特徴の部分集合をブースティングに基づき信頼度付きで選択する。ブースティングのための学習サンプルは葉ノードのクラス $q \in Q^*$ のキー特徴確率分布 $\{p(f_n | q) | f_n \in F\}$ から次のように生成される。まず、サンプル u の局所特徴点数 N_u をある範囲 $[N_{u1}, N_{u2}]$ から一様分布に従って選ぶ。次に、 N_u 個のキー特徴をキー特徴確率分布に従って選んでキー特徴の BoF $H(u) = [h_u(f_1), \dots, h_u(f_{N_F})]$ を作成し、その分布 $D(u)$ を求める。各クラスについて所与の個数のサンプルを生成する。各ブランチノード $b(Q)$ における信頼度付きキー特徴部分集合の選択は次の手順によりなされる。ここで、 Q はこのブランチノードをルートとする部分木の葉ノードに割り当てられるクラスの集合である。ブースティングのための重み付きサンプル集合を $\Phi = \{(u_i, w_i, v_i) | i = 1, \dots, N_\Phi\}$ とする。ここで、 u_i は Q のいずれかのクラスから生成された BoF 分布で表されるサンプル、 w_i はサンプル重み、 $v_i \in \{1, -1\}$ はサンプルのラベルで、 u_i が $b(Q)$ の左子部分木の葉ノードのクラスから生成されたとき 1、右子部分木の葉ノードのクラスから生成されたとき -1 をとる。キー特徴の選択数を T とする。

Step 1 各サンプルの重みを $w_i = 1/N_\Phi$ で初期化する。

Step 2 $t = 1, \dots, T$ に対して、サンプルの重み分布のもとでキー特徴インデクス λ_t とその信頼度 α_t を選択する。いま、 $p_1(f_\lambda)$ 、 $p_2(f_\lambda)$ をそれぞれ左子ノードと右子ノードのキー特徴確率分布の λ 番目の要素の値、 $d_u(\lambda)$ をサンプル u の BoF 分布 $D(u)$ の λ 番目の要素の値とすると、 $\eta(\lambda, u)$ を次のように定義する。

$$\eta(\lambda, u) = \begin{cases} 1 & |p_1(\lambda) - d_u(\lambda)| < |p_2(\lambda) - d_u(\lambda)| \\ 0 & |p_1(\lambda) - d_u(\lambda)| = |p_2(\lambda) - d_u(\lambda)| \\ -1 & |p_1(\lambda) - d_u(\lambda)| > |p_2(\lambda) - d_u(\lambda)| \end{cases} \quad (2)$$

Step 2-1 未選択のキー特徴 f_λ について、すべてのサンプル $(u_i, w_i, v_i) \in \Phi$ に対する誤り重み和 $\epsilon = \sum_{i: v_i \neq \eta(\lambda, u_i)} w_i$ を計算し、誤り重み和を最小とするキー特徴のインデクス λ_t を選ぶ。ここで、誤り重み和を最小とするキー特徴が複数あるときは、 $|p_1(f_\lambda) - p_2(f_\lambda)|$ が最大のキー特徴を選択する。

Step 2-2 誤り重み和から信頼度 $\alpha_t = \frac{1}{2} \log \left(\frac{1-\epsilon}{\epsilon} \right)$ を計算する。ただし、 $\epsilon \geq 0.5$ のときは $\alpha_t = 0$ とする。

Step 2-3 サンプル重みを $w_i = w_i \times \exp(-\alpha_t v_i \eta(\lambda_t, u_i))$ により更新する。

Step 2-4 サンプル重みの和が1になるように正規化する。
Step 3 T 個のキー特徴インデクスと信頼度のペア $\{(\lambda_t, \alpha_t) | t = 1, \dots, T\}$ を記録する。

2.3 情景内物体カテゴリの共起解析

すべての物体クラスがカテゴリラベル付けされると、それら物体クラスのクラス確率から物体カテゴリの出現確率を求めることができる。各情景カテゴリにおける物体カテゴリの共起関係は、その情景カテゴリでの物体カテゴリの出現確率とすべての情景カテゴリでの物体カテゴリの出現確率を用いて次のように計算される。いま、すべての情景カテゴリでの物体カテゴリ c_o の出現確率を $p(c_o)$ 、ある情景カテゴリ c_s における物体カテゴリ c_o の出現確率を $p(c_o|c_s)$ とする。このとき、情景カテゴリ c_s における前景物体カテゴリ $c_f (= c_s)$ と後景物体カテゴリ c_b の共起度 $\omega(c_f, c_b|c_s)$ を

$$\omega(c_f, c_b|c_s) = \begin{cases} \log \frac{p(c_f|c_s)p(c_b|c_s)}{p(c_f)p(c_b)} & \omega(c_f, c_b|c_s) \geq 0 \\ 0 & otherwise \end{cases} \quad (3)$$

で定義する。

2.4 共起とブースト特徴を用いた情景内物体の認識

与えられた情景画像に対して、それに含まれる物体のカテゴリと情景カテゴリを確率潜在コンポーネント木の探索と共起に基づき求める。いま、与えられた情景画像に含まれる物体セグメントの集合を $s_I = \{s_i\}$ 、その数を N_{s_I} 、各物体セグメント s_i の BoF 分布を $D(s_i) = [d_{s_i}(f_1), \dots, d_{s_i}(f_{N_F})]$ とする。

物体カテゴリは次のように求められる。まず、各物体セグメント s_i に対して、その BoF 分布 $D(s_i)$ を用いて各ノードで計算される距離が小さい順にルートからノードを辿ることにより、高々 n_l 個の葉ノードを探索し、 n_c 位までの物体カテゴリを求める。ここで、ブランチノードでの距離は信頼度付きブーストキー特徴を用いて次のように計算される。いま、ブランチノード b のキー特徴インデクスと信頼度のペアの集合を $\{(\lambda_t^b, \alpha_t^b) | t = 1, \dots, T\}$ とする。このとき、すべてのブランチノードの信頼度を用いて、キー特徴 f_n の合成信頼度を

$$\alpha_n^* = \frac{\sum_b \hat{\alpha}_n^b}{\sum_{j=1}^{N_F} \sum_b \hat{\alpha}_j^b}, \quad n = 1, \dots, N_F \quad (4)$$

$$\hat{\alpha}_n^b = \begin{cases} \alpha_t^b & \exists t, n = \lambda_t^b \\ 0 & otherwise \end{cases} \quad (5)$$

と定める。そして、この合成信頼度を用いて、ノード b の左子ノード b^1 と右子ノード b^2 について、合成信頼度付き距離を

$$E(s_i, b^k) = \sum_{n=1}^{N_F} \alpha_n^* \times |d_{s_i}(f_n) - p(f_n|b^k)|, \quad k = 1, 2 \quad (6)$$

で求める。ここで、 $\{p(f_n|b^k)\}$ は子ノード b^k のキー特徴確率分布である。次に、物体セグメント毎に n_c 個求められた物体カテゴリの組合せ $c_{s_I}^k = \{c_{s_i}^k | i = 1, \dots, N_{s_I}\}$ に対して、その組の共起度を次式により求める。

$$\omega(c_{s_I}^k) = \frac{\sum_{s_i} \sum_{s_j (s_j \neq s_i)} \omega(c_{s_i}^k, c_{s_j}^k | c_{s_I}^k)}{N_{s_I}} \quad (7)$$

そして、物体カテゴリの組 $c_{s_I}^k$ に対応する物体クラスの集合と物体セグメント集合 s_I との間で計算される平均距離 $m(c_{s_I}^k)$ とこの共起度 $\omega(c_{s_I}^k)$ に対して、

$$E(c_{s_I}^k) = \gamma \times \delta(\omega(c_{s_I}^k)) - m(c_{s_I}^k) \quad (8)$$

$$\delta(\omega(c_{s_I}^k)) = \begin{cases} 1 & \omega(c_{s_I}^k) > 0 \\ 0 & otherwise \end{cases} \quad (9)$$

を求める。式 (8) で、 γ は共起の影響を定める定数である。最後に、 $E(c_{s_I}^k)$ を最大にする物体カテゴリの組合せ $c_{s_I}^* = \{c_{s_i}^* | i = 1, \dots, N_{s_I}\}$ を物体セグメント集合 s_I に対する物体カテゴリの組として選択する。

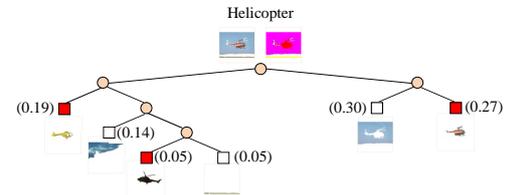
物体カテゴリの組に対して、情景カテゴリを次のように求める。まず、すべての情景カテゴリの中から、その前景物体カテゴリがこれら選択された物体カテゴリのいずれかである情景カテゴリを候補情景カテゴリとして枚挙する。次に、物体セグメントの BoF の和として求められる合成 BoF と情景カテゴリ候補の合成キー特徴確率分布との間の距離を求め、最小の距離を与える情景カテゴリを選択する。このとき、この情景カテゴリと同じラベルの物体が情景画像の前景であると推測される。

3. 実験

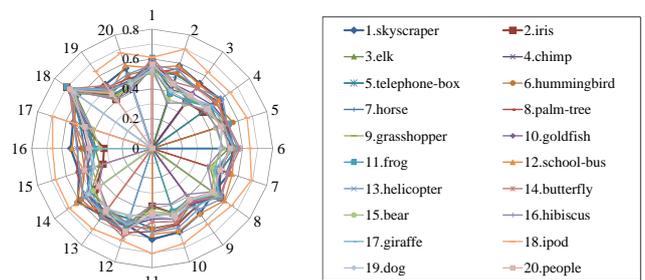
3.1 注意にガイドされた情景の解析実験

注意にガイドされて選択されたセグメント集合からの、潜在クラス数可変の確率潜在コンポーネント解析による情景内物体クラス解析と情景カテゴリの特徴づけの評価を行った。実験には Caltech-256 画像データセットの 20 個のカテゴリの画像を用いた。各カテゴリの画像には、そのカテゴリの物体とそれ以外のカテゴリの物体が含まれる。局所特徴には 128 次元の SIFT 特徴を用い、キー特徴集合のサイズは 438 であった。

図 2 に、情景カテゴリの情景内物体クラス解析の例と 20 個の情景カテゴリの合成キー特徴確率分布間の距離を示す。情景カテゴリ内の物体クラス数は平均 7.55 個、情景カテゴリの合成キー特徴確率分布間の距離は平均 0.51 で、情景カテゴリが合成キー特徴確率分布で特徴づけられることが確かめられた。



(a) 情景内物体クラス解析の例(括弧内の値はクラス確率)



(b) 情景カテゴリの合成キー特徴確率分布間の距離

図 2: 注意にガイドされた情景の解析

3.2 特徴選択と共起に基づく認識実験

確率潜在コンポーネント木を用いた認識におけるブースト特徴と共起の効果の評価を、MSRC ラベル付き画像 DB v2^{*1} に含まれる 429 枚の画像から 16 個の情景カテゴリを構成して、5 分割交差検定により行った。情景カテゴリの画像は前景とし

*1 <http://research.microsoft.com/vision/cambridge/recognition/>

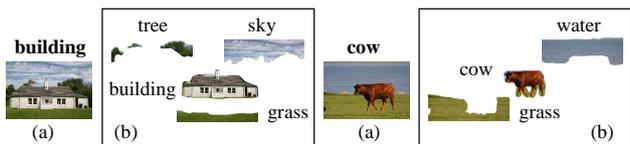


図 3: 16 個の情景カテゴリ (“airplane”, “bicycle”, “bird”, “building”, “car”, “cat”, “chair”, “cow”, “dog”, “grass”, “road”, “sheep”, “sign”, “sky”, “tree”, “water”) の (a) 情景画像と (b) ラベル付き物体セグメントの例

表 1: 物体カテゴリの分類正確度

認識方法	特徴選択	有	無	有	無
	共起利用	有	有	無	無
特徴記述子	DOCS	0.760	0.740	0.742	0.728
	IPGS	0.681	0.674	0.655	0.649

てそのカテゴリの物体セグメントを、背景としてそれ以外のカテゴリの物体セグメントを含む。図 3 に、実験で用いた情景カテゴリの画像と、物体セグメントとそのラベルの例を示す。局所特徴には、疎関心点での 128 次元のグレイ SIFT (Interest Point Grey SIFT, IPGS) と密格子点での 384 次元の反対色 SIFT (Dense Opponent Color SIFT, DOCS) [Sande 10] を用い、IPGS と DOCS 特徴に対するキー特徴集合のサイズはそれぞれ 719 と 720 であった。ブースティングに関して、クラス毎のサンプル数を 500、サンプルの局所特徴点の範囲を $N_{u1} = 50, N_{u2} = 500$, キー特徴選択数の上限 T を 500 に設定した。また、共起に関して、式 (8) の γ を 1.0 に設定した。

IPGS または DOCS 特徴のもとで 16 個の情景カテゴリから生成された物体クラスの総数の平均は 97.6 個 (情景カテゴリあたり 6.1 個) で、これらクラスを葉に持つ分類木の深さの平均は 11.93 であった。また、前景物体カテゴリと平均 2.03 個の後景物体カテゴリとの間に強い共起がみられた。表 1 に、特徴選択有・無、共起利用有・無に対する物体カテゴリの分類正確度を示す。ここで、探索物体カテゴリ数の上限は $n_c = 5$ である。特徴選択、及び共起利用により分類正確度が高くなることが確認された。図 4 に葉ノード探索数 n_l と物体分類正確度の関係を示す。特徴選択及び共起により、少ない探索数で分類正確度が高くなることが確かめられた。また、前景物体、即ち求められた情景カテゴリと同一ラベルの物体の分類正確度は、DOCS 特徴では、特徴選択・共起ありの場合 0.988、なしの場合 0.979、IPGS 特徴では、特徴選択・共起ありの場合もなしの場合も 0.996 と高い性能を示した。

4. おわりに

本論では、注意にガイドされて選択されたセグメント集合からクラス数可変の確率潜在コンポーネント解析により情景内物体カテゴリ解析と情景カテゴリの特徴づけを行う手法、及びブースト特徴を持つ確率潜在コンポーネント木の学習と情景の共起性解析に基づき情景内の物体カテゴリを認識する手法について述べた。本手法では、物体クラス数と分類木の深さが自動的に定まるため、これらを事前にサイズパラメータとして与える必要がなく、また、キー特徴が信頼度付きで選択されるため、所与のキー特徴集合の影響を受けにくくその識別力を増強できる。画像データセットを用いた実験により、注意にガイドされて選択された共起セグメントが情景カテゴリを特徴づける

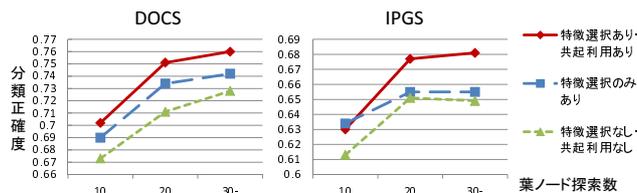


図 4: 葉ノード探索数に対する分類正確度

こと、及びブースト特徴選択と共起の利用が分類正確度を高めることを示して本手法の有用性を確かめた。

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

[Itti 01] Itti, L. and Koch, C.: Computational Modelling of Visual Attention, *Nature Reviews Neuroscience*, Vol.2, pp.194-203 (2001)

[Walther 05] Walther, D., et al.: Selective Visual Attention enables Learning and Recognition of Multiple Objects in Cluttered Scenes, *Computer Vision and Image Understanding*, Vol.100, pp.41-63 (2005)

[Atsumi 08] Atsumi, M.: Attention-based Segmentation on an Image Pyramid Sequence, *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science*, Vol.5259, pp.625-636, Springer-Verlag (2008)

[Liu 11] Liu, T., et al.: Learning to Detect a Salient Object, *IEEE TPAMI*, Vol.33, No.2, pp.353-367 (2011)

[Bosch 06] Bosch, A., et al.: Scene Classification via pLSA, *Proc. of the ECCV*, pp.517-530 (2006)

[Huang 09] Huang, S. and Jin, L.: A PLSA-based Semantic Bag Generator with Application to Natural Scene Classification under Multi-instance Multi-label Learning Framework, *Proc. of the 5th Int. Conf. on Image and Graphics*, pp.331-335 (2009)

[Sivic 08] Sivic, J., et al.: Unsupervised Discovery of Visual Object Class Hierarchies, *Proc. of IEEE CS Conf. on CVPR*, pp.1-8 (2008)

[Galleguillos 08] Galleguillos, C., et al.: Object Categorization using Co-Occurrence, Location and Appearance, *Proc. of IEEE CS Conf. on CVPR* pp.1-8 (2008)

[Choi 12] Choi, M.J., et al.: A Tree-based Context Model for Object Recognition, *IEEE TPAMI*, Vol. 34, No. 2, pp.240-252 (2012)

[Tu 05] Tu, Z.: Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition and Clustering, *Proc. of 10th IEEE Int. Conf. on Computer Vision*, Vol.2, pp.1589-1596 (2005)

[Antenreiter 06] Antenreiter, M., et al.: Visual Classification of Images by Learning Geometric Appearances through Boosting, *Proc. of the Second IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pp.233-243 (2006)

[Atsumi 10] Atsumi, M.: Probabilistic Object Learning through Attention-based Organized Perception, *Proc. of the Joint 5th Int. Conf. on Soft Computing and Intelligent Systems and 11th Int. Symposium on Advanced Intelligent Systems*, pp.304-309 (2010)

[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)

[Sande 10] van de Sande, K. E. A., et al.: Evaluating Color Descriptors for Object and Scene Recognition, *IEEE TPAMI*, Vol.32, pp.1582-1596 (2010)