

# 前景・後景コンテキストからの ラベル付き物体カテゴリ木学習に基づく共起物体認識

## Co-occurrence Object Recognition based on Learning a Labeled Object Category Tree from Foreground/Background Contexts

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 probabilistic learning method that is named a contextual probabilistic latent component tree for object and scene categorization. In this method, object classes are obtained by clustering a set of object segments extracted from scene images of each scene category and their categorical co-occurrence relations in scene categories are embedded in the probabilistic latent component tree that is generated as a classification tree of all the object classes of all the scene categories. Through experiments by using images of plural categories in an image database, it is shown that the co-occurrence relation of object categories in scene categories improves performance for object and scene recognition.

### 1. はじめに

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いる。前者は物体カテゴリ間の共起関係により表現され、後者は物体の見え特徴の類似木により表現される。本論では、物体と情景の認識のために、情景毎の物体の共起関係を埋め込んだ物体の見えの分類木を学習する問題を扱う。本問題では、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。

本論では、この問題に対して、コンテキストを伴う確率潜在コンポーネント木 (Contextual Probabilistic Latent Component Tree, C-PLCT) と名付ける確率的学習手法と C-PLCT を用いた物体と情景の認識手法について述べる。C-PLCT 学習では、各情景カテゴリの情景画像から抽出された物体セグメント集合に対して、まず、物体クラス集合が潜在クラス数可変の確率潜在コンポーネント解析 (Probabilistic Latent Component Analysis with the Variable number of classes, V-PLCA) により物体セグメント集合をクラスタリングすることにより求められる [Atsumi 10]。次に、すべての情景カテゴリのすべての物体クラス集合に対して、それらの分類木として確率潜在コンポーネント木 (Probabilistic Latent Component Tree, PLCT) が物体クラス間の類似性に基づき生成され、その葉ノードに位置する物体クラスがそれら物体クラスの代表インスタンスに対して教師信号として与えられるカテゴリ名を用いて半教師付きでラベル付けされる [Atsumi 11]。そして最後に、各情景カテゴリの物体カテゴリ間の共起関係がラベル付けされた物体クラスから求められる。物体と情景の認識では、複数の物体セグメントを含む情景画像に対して、C-PLCT を用いて物体カテゴリの集合とそれらの複合としての情景カテゴリが識別される。

関連する研究として、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た手法 [Sivic 05, Bosch 06] やカテゴリゼーションを階層化した手法が提案されている。Sivic ら [Sivic 08] は階層 LDA を用いて物体カテゴリ階層を生成しているが、階層の深さはあらかじめ決められている。コンテキストが物体認識のあいまいさを改善する効果 [Bar 04] に着目し、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 10] も提案されている。我々の提案手法の既存手法との主な違いは、情景カテゴリ毎の物体カテゴリの共起関係を物体カテゴリの見えの分類木に埋め込んでいる点と、それに基づいて物体と情景カテゴリを同時に認識する点である。

以下、2章で提案手法、3章で実験結果について述べる。

### 2. 物体と情景の学習と認識のモデル

情景カテゴリの集合を  $C$ 、情景カテゴリの数を  $N_c$  とする。情景カテゴリ  $c \in C$  を、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情景画像の集合とする。情景カテゴリ  $c$  の情景画像  $i$  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  $c$  の任意の情景画像から分節された物体セグメントの集合を  $S_c$ 、 $N_{c_s}$  を  $S_c$  の物体セグメントの数とする。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 [Csurka 04]。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。

情景カテゴリ集合  $C$  の各情景カテゴリ  $c \in C$  に対して、その情景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、C-PLCT の学習問題は、各  $H_c$  から物体クラスの集合  $Q_c$  を求め、全物体クラスの集合  $\cup_{c \in C} Q_c$  からそれらクラスを葉に持ちかつそれらが物体カテゴリ名でラベル付けされた分類木を求め、そして、各情景カテゴリ  $c \in C$  の前景物体カテゴリと後景物体カテゴリの共起関係を求める問題である。C-PLCT を用いた認識問題は、複数の物体セグメントを含む情景画像が与えられて、それら物体セグメントの各々の物体カテゴリとそれらの複合としての情景カテゴリを類似性と共起関係を用いて同時に推定する問題である。

## 2.1 情景の確率潜在コンポーネント解析

各情景カテゴリ  $c \in C$  の物体カテゴリを表すクラス集合  $Q_c = \{q_{c,r} | r = 1, \dots, N_{c_q}\}$  を潜在クラス数可変の確率潜在コンポーネント解析 (V-PLCA) により求める。このために、各情景カテゴリ  $c \in C$  の BoF 集合  $H_c$  に対して、対数尤度

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

を最大にする  $p(s_{c,i_j}, f_n) = \sum_r p(q_{c,r}) p(s_{c,i_j} | q_{c,r}) p(f_n | q_{c,r})$ , 即ちクラス確率  $\{p(q_{c,r}) | q_{c,r} \in Q_c\}$ , インスタンスの条件付き確率  $\{p(s_{c,i_j} | q_{c,r}) | s_{c,i_j} \in S_c, q_{c,r} \in Q_c\}$ , キー特徴の条件付き確率分布  $\{p(f_n | q_{c,r}) | f_n \in F, q_{c,r} \in Q_c\}$ , 及び潜在クラスの数  $N_{c_q}$  を推定する。ここで、クラス確率は情景カテゴリ内の物体カテゴリの存在比率、インスタンスの条件付き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、キー特徴の条件付き確率分布はクラスが表す物体カテゴリの特徴を与える。

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 [Atsumi 10, Atsumi 11]. 即ち、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ばらつき度

$$\delta_{c,r} = \sum_{s_{c,i_j}} \left( \sum_{f_n} |p(f_n | q_{c,r}) - D(s_{c,i_j}, f_n)| \right) \times p(s_{c,i_j} | q_{c,r}) \quad (2)$$

ここで、

$$D(s_{c,i_j}, f_n) = \frac{h_{c,i_j}(f_n)}{\sum_{f_{n'}} h_{c,i_j}(f_{n'})} \quad (3)$$

を求めて、ばらつき度が最大値のクラスを 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。

## 2.2 確率潜在コンポーネント木による物体カテゴリ分類

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 (PLCT) をそれらクラスを葉を持つ分類木として生成する。PLCT は類似の物体クラスが近くの葉に配置される 2 分木で、類似性はそれらクラスのキー特徴の条件付き確率分布とクラス確率を用いて計算される。

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。ここで、PLCT のルートノードに対しては  $Q^0 = Q^*$  である。このとき、まず、各キー特徴  $f_n \in F$  について、 $Q^0$  をそれに含まれるクラスのキー特徴  $f_n$  の確率値に応じて、2 つの部分集合  $Q_{f_n}^1 = \{q_{c,r} | p(f_n | q_{c,r}) \leq \epsilon, q_{c,r} \in Q^0\}$  と  $Q_{f_n}^2 = \{q_{c,r} | p(f_n | q_{c,r}) > \epsilon, q_{c,r} \in Q^0\}$  に分割する。ここで、 $\epsilon$  は 0 または非常に小さい正の整数で、デフォルトでは 0 とする。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

$$D_{f_n} = \sum_{q_{c,r} \in Q_{f_n}^1} p(q_{c,r}) \left( \sum_{f_{n'} \in F} p(f_{n'} | q_{c,r}) \log \frac{p(f_{n'} | q_{c,r})}{\mu_{Q_{f_n}^1}(f_{n'})} \right) \quad (4)$$

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

を計算する。最後に、すべてのキー特徴  $f_n \in F$  に対して、 $D_{f_n}$  の最小値を与えるクラス集合分割  $Q^1, Q^2$  を求める。そして、それら部分クラス集合  $Q^1, Q^2$  の各々に対して、クラス数が 1 より大きい場合はそれぞれブランチノード  $B(Q^1)$  と  $B(Q^2)$  を生成し、ブランチノード  $B(Q^0)$  の子ノードとする。一方、クラス数が 1 の場合は、そのクラスを持つ葉ノードを生成し、ブランチノード  $B(Q^0)$  の子ノードとする。以上の処理を、PLCT のルートノード  $B(Q^*)$  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、キー特徴の条件付き確率分布、及び物体セグメントの条件付きインスタンス確率が保持される。ただし、葉ノードのクラス確率は、そのクラスのクラス確率を情景カテゴリ数で割った値  $p(q_{c,r})/N_c$  とする。

葉の物体クラスのカテゴリラベルは、いくつかの物体クラスの条件付き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 [Atsumi 11]. このラベル付けにより情景カテゴリの各物体クラスが前景物体カテゴリを表すか後景物体カテゴリを表すかが定まる。情景カテゴリの特徴は、その前景及び後景物体カテゴリを表すクラスのキー特徴の条件付き確率分布を合成することにより表される。いま、 $Q_c^f(\theta_f) = \{q_{c,r} | q_{c,r} \in Q_c, p(q_{c,r}) \geq \theta_f\}$  と  $Q_c^b(\theta_b) = \{q_{c,r} | q_{c,r} \in Q_c, p(q_{c,r}) \geq \theta_b\}$  をそれぞれ、与えられたクラス確率の閾値  $\theta_f$  と  $\theta_b$  に対する  $Q_c^f$  と  $Q_c^b$  の部分集合とする。このとき、情景カテゴリ  $c$  の合成キー特徴確率分布は、各  $\forall f_n \in F$  に対して、

$$p(f_n | Q_c^f(\theta_f), Q_c^b(\theta_b)) = \sum_{q_{c,r} \in Q_c^f(\theta_f) \cup Q_c^b(\theta_b)} \lambda(q_{c,r}) \times p(f_n | q_{c,r}) \quad (5)$$

ここで、

$$\lambda(q_{c,r}) = \frac{p(q_{c,r})}{\sum_{q_{c,r'} \in Q_c^f(\theta_f) \cup Q_c^b(\theta_b)} p(q_{c,r'})} \quad (6)$$

で求められる。

## 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

$$\omega(c_f, c_b | c_s) = \begin{cases} \log \frac{p(c_f | c_s) p(c_b | c_s)}{p(c_f) p(c_b)} & \omega(c_f, c_b | c_s) \geq \epsilon_\omega \\ \epsilon_\omega & \text{otherwise} \end{cases} \quad (7)$$

で定義する。ここで、 $\epsilon_\omega$  は共起度の下限を定める定数でデフォルトで 0 とする。

## 2.4 共起性に基づく物体と情景の認識

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のカテゴリを、C-PLCT を用いて類似性と共起性に基

づき認識する2つの手法-「物体主導情景認識 (Object-to scene recognition)」と「物体共起主導情景認識 (Co-object-to-scene recognition)」-について述べる。

物体主導情景認識では、まず、情景画像内の各物体セグメントに対して、物体セグメントの BoF と C-PLCT の葉クラスのキー特徴条件付き確率分布との間の距離を計算して、最小の距離を与える物体カテゴリを選択する。次に、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。そして最後に、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。

物体共起主導情景認識では、上記の物体セグメントのカテゴリの認識に物体カテゴリ間の共起関係を利用する。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  $N_{s_I}$  とする。このとき、まず、各物体セグメント  $s_i$  に対して、それとの距離が小さい  $n$  位までの物体カテゴリを枚挙し、次に、物体セグメント集合に対する物体カテゴリの組  $c_{s_I}^k = \{c_{s_i}^k | i = 1, \dots, N_{s_I}\}$  に対して、その組の共起度を次式により求める。

$$\omega(c_{s_I}^k) = \frac{\sum_{s_i} \sum_{s_j (s_j \neq s_i)} \omega(c_{s_i}^k, c_{s_j}^k | c_{s_i}^k)}{N_{s_I}} \quad (8)$$

そして、物体カテゴリの組  $c_{s_I}^k$  を与える物体クラスの組と物体セグメント集合  $s_I$  との距離の平均  $d(c_{s_I}^k)$  とこの共起度  $\omega(c_{s_I}^k)$  に対して、

$$E(c_{s_I}^k) = \gamma \times \delta(\omega(c_{s_I}^k)) - d(c_{s_I}^k) \quad (9)$$

ここで、

$$\delta(\omega(c_{s_I}^k)) = \begin{cases} 1 & \omega(c_{s_I}^k) > 0 \\ 0 & otherwise \end{cases} \quad (10)$$

を求める。式 (9) で、 $\gamma$  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  $c_{s_I}^0 = \{c_{s_i}^0 | i = 1, \dots, N_{s_I}\}$  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。

### 3. 実験

#### 3.1 実験枠組

物体と情景の学習と認識の実験を、MSRC ラベル付き画像 DB v2<sup>\*1</sup> 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行った。各情景カテゴリは約 27 枚の情景画像を含み、各画像は前景としてそのカテゴリの物体セグメントを、後景としてそれ以外のいくつかのカテゴリの物体セグメントを含む。図 1 に、実験で用いたいくつかのカテゴリの画像と、それら画像から分節されたセグメントとそのラベルを示す。これら情景カテゴリの画像は、ほぼ同じサイズに 5 分割されて交差検定に用いられた。実験で用いた主なパラメータは次のとおりである。V-PLCA のクラス数決定過程におけるばらつき度とクラス確率の閾値はそれぞれ 1.0 と 0.2 に、式 (5) と (6) における閾値  $\theta_f$  と  $\theta_b$  は 0.1 に、物体共起主導情景認識の式 (9) における  $\gamma$  は 1.0 に設定された。

疎関心点での 128 次元のグレイ SIFT [Lowe 04] と密格子点での 384 次元の反対色 SIFT [Sande 10] の 2 つのタイプの局

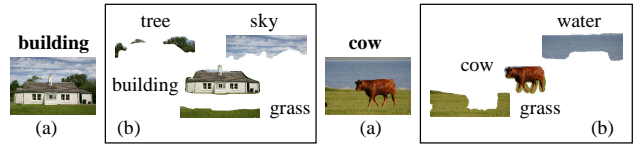


図 1: Examples of (a) scene images and (b) object segments with labels. Scene images and object segments of 16 categories (“airplane”, “bicycle”, “bird”, “building”, “car”, “cat”, “chair”, “cow”, “dog”, “grass”, “road”, “sheep”, “sign”, “sky”, “tree”, “water”) were used in experiments.

所特徴量が実験で用いられた。これら 2 つの特徴量をそれぞれ IPGS (Interest Point Grey SIFT) と DOCS (Dense Opponent Color SIFT) と略記する。キー特徴のコードブックは K-tree 手法 [Shlomo 00] により求められ、IPGS と DOCS 特徴に対するコードブックサイズはそれぞれ 719 と 720 であった。

#### 3.2 実験結果

IPGS または DOCS 特徴のもとで 16 個の情景カテゴリから V-PLCA により生成された物体クラスの総数の平均は 97.6 個で、これらクラスを葉に持つ PLCT の深さの平均は 11.93 であった。表 1 に、DOCS 特徴を用いた C-PLCT の情景カテゴリにおける前景・後景物体カテゴリ間の平均共起マトリクスを示す。また、図 2 に、情景カテゴリの前景・後景物体クラスとそれらの構成比の例を示す。前景物体カテゴリと平均で 2.5 個の後景物体カテゴリとの間に強い共起がみられた。

表 2 に、2 つの提案認識手法による物体と情景の認識率の平均を示す。物体共起主導情景認識手法 (C-O-S) が物体主導情景認識手法 (O-S) より高い認識率を達成した。また、DOCS 特徴を用いるの方が IPGS 特徴を用いるよりも高い認識率を達成した。

### 4. おわりに

本論では、物体と情景の学習と認識のためのコンテキストを伴う確率潜在コンポーネント木 (C-PLCT) 手法について述べた。V-PLCA では、情景カテゴリの物体クラス数は、学習データに依存して定められ、あらかじめ与える必要はない。また、PLCT では、分類木の深さは、V-PLCA で生成される物体クラス集合に依存して定められ、あらかじめ与える必要はない。これら特徴により、本手法はサイズに関するパラメータを調整する必要なく様々な学習データセットや特徴量に適応が可能である。

本手法は、物体カテゴリと情景カテゴリの両方を同時に学習、及び認識することができる。DOCS 特徴を用いた物体共起主導情景認識手法は、情景に関して 0.809、物体に関して 0.740 の

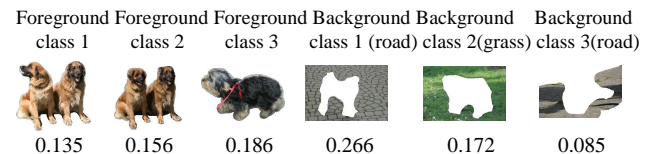


図 2: An example of object category composition of a scene category “dog”. Images are representative segments whose instance probabilities are maximum. Class probabilities which represent composition ratio are shown under those segment images.

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

表 1: Co-occurrence matrix between foreground and background object categories

object scene	road	grass	sheep	cow	dog	bird	sky	airplane	car	sign	building	tree	water	bicycle	cat	chair
road	-	0.0	0.0	0.0	0.0	0.0	1.20	0.0	0.0	0.0	1.75	1.41	0.0	0.0	0.0	0.0
grass	0.0	-	1.67	3.18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
sheep	0.0	3.60	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
cow	0.0	3.15	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.66	0.0	0.0	0.0
dog	3.75	2.79	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
bird	0.0	2.85	0.0	0.0	0.0	-	2.63	0.0	0.0	0.0	0.0	0.0	2.52	0.0	0.0	0.0
sky	0.90	0.55	0.0	0.0	0.0	0.0	-	1.63	0.0	0.0	1.18	1.75	0.0	0.0	0.0	0.0
airplane	2.71	2.56	0.0	0.0	0.0	0.0	3.55	-	0.0	0.0	0.81	0.0	0.0	0.0	0.0	0.0
car	3.78	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	2.51	1.72	0.0	0.0	0.0	0.0
sign	0.0	0.0	0.0	0.0	0.0	0.0	0.51	0.0	0.0	-	3.24	3.44	0.0	0.0	0.0	0.0
building	0.0	0.77	0.0	0.0	0.0	0.0	1.78	0.0	0.0	0.0	-	2.00	0.0	0.0	0.0	0.0
tree	0.0	0.24	0.0	0.0	0.0	0.0	3.07	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0
water	0.0	0.0	0.0	0.0	0.0	0.0	3.03	0.0	0.0	0.0	2.91	0.0	-	0.0	0.0	0.0
bicycle	2.91	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.59	0.0	0.0	-	0.0	0.0
cat	4.22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0
chair	0.82	2.62	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-

表 2: Recognition rates of objects and scenes. In recognition methods, “O-S” represents the object-to-scene recognition method and “C-O-S” represents the co-object-to-scene recognition method. The number in a parenthesis indicates the number of top categories shortlisted for object segments. The object recognition accuracy is a rate in which a selected object category is the same with its ground truth and the scene recognition accuracy is a rate in which a selected scene category is one of ground truth of objects in the scene.

Feature type	DOCS			IPGS		
Recognition method	O-S	C-O-S(5)	C-O-S(10)	O-S	C-O-S(5)	C-O-S(10)
Object recognition accuracy	0.728	0.740	0.737	0.649	0.674	0.683
Scene recognition accuracy	0.807	0.809	0.809	0.676	0.690	0.695

認識率を達成した。一般に、認識性能は、学習・認識手法のみでなく、特徴量のコーディング・プーリング、及び学習データセットに依存する [Boureau 10]。本実験結果より、本手法が、単純な SIFT 特徴を用いる場合の既存手法と比べて、同等以上の高い性能を示すことが確かめられた [Sande 10, Boureau 10]。

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

- [Atsumi 10] Atsumi, M: Learning Visual Object Categories and Their Composition Based on a Probabilistic Latent Variable Model, Neural Information Processing: Theory and Algorithms (Proceedings of ICONIP 2010, Part I), Lecture Notes in Computer Science, Vol.6443, pp.247-254, Springer-Verlag (2010)
- [Atsumi 11] Atsumi, M: Visual Category Learning based on Probabilistic Latent Component Models with Semi-supervised Labeling, Proc. of 2nd Annual Int. Conf. on Advanced Topics in Artificial Intelligence, pp.50-55 (2011)
- [Sivic 05] Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A. and Freeman, W. T.: Discovering Objects and their Location in Images, Proc. of IEEE Int. Conf. on Computer Vision, pp.370-377 (2005)
- [Bosch 06] Bosch, A., Zisserman, A. and Munoz, X.: Scene Classification via pLSA, Proc. of the European Conf. on Computer Vision, pp.517-530 (2006)
- [Sivic 08] Sivic, J., Russell, B. C., Zisserman, A., Freeman, W. T. and Efros, A. A.: Unsupervised Discovery of Visual Object Class Hierarchies, Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, pp.1-8 (2008)
- [Bar 04] Bar, M.: Visual Objects in Context, Nature Reviews Neuroscience, Vol.5, pp.617-629 (2004)
- [Galleguillos 08] Galleguillos, C., Rabinovich, A. and Belongie, S.: Object Categorization using Co-Occurrence, Location and Appearance, Proc. of IEEE CS Conf. on CVPR pp.1-8 (2008)
- [Choi 10] Choi, M. J., Lim, J. J., Torralba, A. and Willsky, A. S.: Exploiting Hierarchical Context on a Large Database of Object Categories, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (2010)
- [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)
- [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)
- [Sande 10] Van de Sande, K. E. A., Gevers, T. and Snoek, C. G. M.: Evaluating Color Descriptors for Object and Scene Recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.32, pp.1582-1596 (2010)
- [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)
- [Boureau 10] Boureau, Y. L., Bach, F., LeCun, Y. and Ponce, J.: Learning Mid-Level Features For Recognition, Proc. of 2010 IEEE Conf. on Computer Vision and Pattern Recognition, pp.2559-2566 (2010)