一般物体認識におけるマルチモーダル星座モデルの提案

神谷　保徳† 　高橋　友和†† 　井手　一郎† 　村瀬　洋†

† 名古屋大学大学院情報科学研究科，〒 464-8603 名古屋市千種区不老町
†† 岐阜聖徳学園大学経済情報学部，〒 500-8288 岐阜市中鶉 1-38
E-mail: †kamiya@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

あらまし　画像の撮影条件や対象とする物体の種類を制限せずに物体の属するカテゴリを認識する一般物体認識は，現在注目されている物体認識の研究分野の一つである．この分野には，現在主流になりつつある手法として Bag of Features (BoF) が存在し，また星座モデルも提案されている．星座モデルには，(a) 分類候補カテゴリの追加や変更が容易，(b) BoF に比べ記述精度が高い，(c) BoF では無視する位置とスケールの情報を適切に利用可能，といったメリットがある．しかし，(1) 本質的にユニモーダル（単峰性）なモデルであるため，カテゴリ中の物体が見た目の大きく異なる複数の種類に分かれる場合には記述性能が低い，(2) 星座モデルを表現する確率分布関数の計算に多量の時間が掛かる，といった欠点が存在する．本稿では，これら欠点を解決する“ マルチモーダル星座モデル ”を提案する．実験では，BoF を用いた手法と比較し提案手法の有効性を確認する．

キーワード　星座モデル，マルチモーダル化，高速化，一般物体認識，EM アルゴリズム

Proposal of Multimodal Constellation Model for Generic Object Recognition

Yasunori KAMIYA†, Tomokazu TAKAHASHI††, Ichiro IDE†, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
†† Faculty of Economics and Information, Gifu Shotoku Gakuen University
1-38, Nakauzura, Gifu, 500-8288, Japan
E-mail: †kamiya@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

Abstract　Object category recognition in various appearances is one of the most challenging object recognition research fields. The major approach to solve the task is using the Bag of Features (BoF). The constellation model is another approach that has the following advantages: (a) Adding and changing the candidate categories is easy; (b) Its description accuracy is higher than BoF; (c) Position and scale information, which are ignored by BoF, can be used effectively. On the other hand, this model has two weak points: (1) It is essentially an unimodal model that is unsuitable for categories with many types of appearances. (2) The probability function that represents the constellation model takes a long time to calculate. In this paper we propose “Multimodal Constellation Model” to solve the two weak points of the constellation model. Experimental results showed the effectivity of the proposed model by comparison to methods using BoF.

Key words　Constellation model, Multimodalization, Speed-up, Generic object recognition, EM algorithm.

1. はじめに

我々人間は，車，椅子，焼そばなどの，物体が属するカテゴリ (一般名称) を，物体の向き，距離，明るさ，背景の違いなどの状況変化に因らずに認識することができる．しかしコンピュータが画像から同様の事を行うことは，現在の認識技術では困難である．カテゴリ中には，物体自体の色，形の違いや状況変化によって様々に見た目が変化する物体が含まれるため，

特徴抽出，モデルの構築，学習データセットの構築が困難となるためである．このような物体認識は一般物体認識と呼ばれ，物体認識の分野における困難な課題の一つとされている [1].

一般物体認識では，画像の特徴的な部分領域を局所特徴として用いる Part-based アプローチが広く用いられている．部分的な領域に着目することで，物体のアピアランスの変化に柔軟に対応することができる．代表的なモデルとして，Bag of Features（BoF）[2] を用いた手法と，Fergus の星座モデル（constellation model）[3] がある．BoF は自然言語処理における Bag of Words モデルのアナロジーであり，BoF を用いた手法として，SVM 等の分類器を用いたもの（例：[4] [5] [6]）と，probabilistic Latent Semantic Analysis (pLSA)，Latent Dirichlet Allocation (LDA)，Hierarchical Dirichlet Processes (HDP) などの文章解析手法を用いたもの（例：[7] [8] [9]）がある．

星座モデルは，対象カテゴリを確率モデルとして記述する．確率モデルは，ターゲットカテゴリ内の物体に共通する部位の見た目とその位置関係を記述する．記述する部位は５～７が想定される．詳細は 2.1 章で述べる．

星座モデルには，利点として以下の (a)(b)(c) が存在する．

(a)　処理対象カテゴリの追加や変更が容易：この研究分野では，認識手法はしばしば生成モデルと判別モデル [10] に分類される．この利点 (a) は，星座モデルが生成モデルである事に起因する．生成モデルは，処理対象カテゴリをそれぞれ個別にモデル化する．従ってカテゴリ追加時の学習処理は追加するカテゴリのみでよく，処理対象となるカテゴリの変更なら処理に使用するモデルの変更のみでよく学習処理は必要ない．その一方，判別モデルは全ての処理対象カテゴリを区別する決定境界として学習される．そのため，新しいカテゴリの追加や，処理対象カテゴリの変更の度に再学習が必要となる．ただし判別モデルの分類性能は生成モデルを上回ると一般的に言われている．

(b)　連続値表現であるため記述精度が高い：BoF によるカテゴリの記述は，それぞれの codeword に対する局所特徴の個数から成るヒストグラムによる離散表現である．それに対して星座モデルは，確率分布関数による連続値表現であるため，記述精度が BoF よりも高い．

(c)　位置とスケールの情報を適切に利用可能：BoF は煩雑な位置関係の記述を回避するため，局所特徴の位置情報を無視する．対して星座モデルでは，大まかな位置関係を確率分布関数で表現し，カテゴリを記述する情報として用いる．

しかしながら星座モデルには，欠点として以下の (1)(2) が存在する．

(1)　本質的にユニモーダル（単峰性）なモデルであり，そのため，カテゴリ中の物体の見た目が，大きく異なった複数の種類に分かれる場合は記述精度が低い．

(2)　星座モデルを表現する確率分布関数の計算に多量の時間が掛かる．

本稿では，欠点 (1)(2) を改善するため“ マルチモーダル星座モデル ”を提案する．まず欠点 (1) に対して，モデルをマルチモーダル（多峰性）に拡張する．ユニモーダルなモデルでは，複数の種類の見た目を一括で記述しなければならなかったが，マルチモーダルなモデルに拡張することで，複数のそれぞれの種類の見た目を個別に記述可能にし，カテゴリの記述精度の向上を図る．これは，単一ガウス分布による記述から混合ガウス分布による記述への拡張を，局所特徴表現において行ったことに等しい．また欠点 (2) に対して，処理の高速化を行う．

星座モデルの利点 (b)(c) は，他の論文では殆ど述べられていない．従って，本手法の評価実験に加え，4.5 章において定量的な検証を行う．

本稿では，実験で用いるタスクを，通常用いられているデータセットに標準的に設定されているタスクから少し変更した．一般物体認識では，共通のデータセットとタスクを用いた際の正答率を用いて，他の識別手法との比較を行うことが多い．通常識別手法を提案する場合，局所特徴は提案事項には含まれず，実験では適当な局所特徴が使用される．しかし比較される正答率は，実験で使用する局所特徴の違いの影響も受けるため，この比較は正確とは言えない．またたとえ局所特徴の種類が同じでも，局所特徴を用いる識別手法では，局所特徴の検出個数の違いや検出位置の少しの違いにより，正答率が変化する．それにもかかわらず，様々な論文で，重要な比較のように扱われている．この比較には，比較対象となった識別手法の性能について誤った解釈を読者に持たせる危険性がある．本稿では，このような比較から逃れるため，通常のタスクを少し変更し，その正答率を示した．なお，変更したタスクは，タスクの種類としては画像の適切なカテゴリへの分類処理になる．

以降，2. 章でマルチモーダル星座モデルについて述べる．3. 章ではマルチモーダル星座モデルを用いた分類処理について説明する．4. 章で実験について述べ，5. 章で本稿をまとめる．

2.　マルチモーダル星座モデル

まず，Fergus の星座モデルについて述べる．次に星座モデルのマルチモーダル化と，高速化の工夫について説明し，最後にモデルパラメータの推定方法について述べる．

2.1　Fergus の星座モデル [3]

カテゴリに共通する物体の複数の部位に着目し記述する．各部位とその位置関係はガウス分布で表現される．

確率モデルの式は以下となる．

$$
\begin{aligned}
p(I|\Theta) &= \sum_{\mathbf{h}\in H} p(A,X,S,\mathbf{h}|\Theta) \\
&= \sum_{\mathbf{h}\in H} p(A|\mathbf{h},\theta_A)p(X|\mathbf{h},\theta_X)p(S|\mathbf{h},\theta_S)p(\mathbf{h}|\theta_{other}).
\end{aligned}
$$

ここで，I は入力画像，Θ はモデルパラメータである．I は局所特徴の集合として表現される．局所特徴は，見た目，位置，スケール（局所特徴領域の大きさ）の特徴ベクトルを保持する．A は見た目の特徴ベクトルの集合，X は位置の特徴ベクトルの集合，S はスケールの特徴ベクトルの集合である．また，ハイパーパラメータとして部位数 R がある．$\mathbf{h}$ は，画像 I から得られた全ての局所特徴を，モデルが表現する各部位に割り当てる割当て方の一つを表現するベクトルであり，H は，割当て方の全ての組み合わせの集合である．$\sum_{\mathbf{h}\in H}$ により，画像から得られた局所特徴とモデルが表現する部位との割り当ての組み合

わせが網羅されている．$p(A|\mathbf{h},\theta_A)$ は R 個のガウス分布の積として表現される．$p(X|\mathbf{h},\theta_X)$ は各部位の x,y 座標の組を $2R$ 次元の一つのガウス分布として表現される．また，$p(S|\mathbf{h},\theta_S)$ も R 個のガウス分布の積として表現される．詳細は [3] を参照してほしい．

画像から得られた局所特徴とモデルが表現する部位との割り当てを網羅的に計算する部分 $\left(\sum_{\mathbf{h}\in H}\right)$ は和の表現となっているが，対象カテゴリを表現する部分 $(p(A,X,S,\mathbf{h}|\Theta))$ が実質ガウス分布の積であるため，Fergus の星座モデルはユニモーダル（単峰性）である．

2.2 星座モデルのマルチモーダル化

本稿で提案するマルチモーダル星座モデルを以下の様に定義する．

$$p_m(I|\Theta) = \sum_k^K \left\{ \prod_l^L G(\mathbf{x}_l|\theta_{k,\hat{r}_{k,l}}) \right\} \cdot \pi_k$$

$$= \sum_k^K \left\{ \prod_l^L G(\mathbf{A}_l|\theta^{(A)}_{k,\hat{r}_{k,l}}) G(\mathbf{X}_l|\theta^{(X)}_{k,\hat{r}_{k,l}}) G(\mathbf{S}_l|\theta^{(S)}_{k,\hat{r}_{k,l}}) \right\} \cdot \pi_k$$

$$\hat{r}_{k,l} = \arg\max_r G(\mathbf{x}_l|\theta_{k,r}).$$

ここで，K はモデルの構成要素数を表し，この値が 2 以上の場合，モデルはマルチモーダルとなる．k は構成要素のインデックスである．L は画像 I から得られた局所特徴の数，$G(\,)$ はガウス分布を表す．また，$\Theta=\{\theta_{k,r}\,,\,\pi_k\}$，$\theta=\{\boldsymbol{\mu},\boldsymbol{\Sigma}\}$，$I=\{\mathbf{x}_l\}$，$\mathbf{x}=(\mathbf{A},\mathbf{X},\mathbf{S})$ である．$\theta_{k,r}$ は，構成要素 k 中の部位 r のガウス分布のパラメータを，$\mathbf{x}_l$ は l 番目の局所特徴の特徴ベクトルを表す．$\mathbf{A},\mathbf{X},\mathbf{S}$ はそれぞれ，見た目，位置，スケールの特徴ベクトルであり，$\mathbf{x}$ のサブベクトルである．π_k は，各構成要素 k の存在確率である．$\hat{r}_{k,l}$ は，構成要素 k における，局所特徴 l に最も類似する部位のインデックスである．また，式中には現れていないが，この他に部位数 R がハイパーパラメータとして存在する．

提案モデルは，マルチモーダルであることや後述する高速化の工夫が行われていることに加え，Fergus の星座モデルに比べシンプルであるという特徴がある．モデルの構成要素一つは，純粋に，R 個のガウス分布のみで構成できるため，Fergus の星座モデルに比べ実装が容易である．

2.3 高速化の工夫

Fergus の星座モデルを表現する確率分布関数の計算には，計算量が多く掛かり，特にモデルパラメータの学習には非常に長い時間を要する．この事は，星座モデルのマルチモーダル化を困難にする．なぜならマルチモーダル化によりパラメータが増加し，現実的な時間内での学習終了が不可能になるためである．そこで，高速化の工夫が必要となる．ここでは本手法において行った，二点の工夫について述べる．

[行列計算の簡略化] モデル中の全ての共分散行列 Σ について，非対角要素を省略した．これにより，ガウス分布の計算の際に必要な，$(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ や $|\boldsymbol{\Sigma}|$ の計算量が大幅に減少する．計算量は，$D\times D$ 行列とすると，$O(D^3)$ から，$O(D)$ になる．

具体的には，$\boldsymbol{\Sigma}$ を，対角成分が σ_d^2 の対角行列とすると，

$$(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_d^D \frac{1}{\sigma_d^2}(x_d-\mu_d)^2$$

$$|\boldsymbol{\Sigma}| = \prod_d^D \sigma_d^2$$

となる．

多次元ガウス分布を用いた，物体カテゴリの確率分布の記述では，確率分布の形状を決定するパラメータは，平均，Σ の対角要素，Σ の非対角要素，の三つに分けることができる．平均と Σ の対角要素は分布毎に異なる値を持つ．しかし Σ の非対角要素は，各特徴変数間の相関が低いほど 0 に近くなり，変数間に相関がある場合は何らかの値を持つ．ここで行った Σ の非対角要素の省略は，変数間に相関がある場合は，分布形状の記述精度を低下させる．ただし，この記述精度の低下は，マルチモーダル化による記述性能向上により補われる．

[$\Sigma_{h\in H}$ の $\prod_l^L$ と $\arg\max_r$ への変更] Fergus の星座モデルにおける，画像から得られた局所特徴とモデルの部位の割当てを網羅的に計算する部分 $\Sigma_{h\in H}$ では，局所特徴の数を L，部位の数を R とすると，単純には $p(A,X,S,\mathbf{h}|\Theta)$ の計算が $O(L^R)$ 回行われることになる．実際は，A*アルゴリズムなどの高速演算手法を用いて効率化しているため，計算回数はかなり少なくなるが，それでも全体の計算量は大きい．提案手法ではこの部分を $\prod_l^L$ と $\arg\max_r$ へ変更する．それにより計算回数は $O(LR)$ となる．なお，この修正は [11] を参考にして行った．この論文では，定点カメラからの車両画像を車両の種類に分類するタスクを対象として星座モデルの修正を行っている．この修正の中の計算量の削減に関する部分を参考にした．

ここで，これらの記述表現を比較し，Fergus の星座モデルの記述能力と本手法の記述能力がほぼ同等であるということを説明する．まず，各手法の確率モデルの意味を，計算手順と共に説明する．Fergus の星座モデルは，各部位に対して対応する局所特徴を網羅的に検査，各検査時の，部位と局所特徴の割当てにおける確率を計算し，その和として最終的な確率を計算する．対応する局所特徴の網羅的な検査は，$\Sigma_{h\in H}$ により行われる．それに対し，本手法では，全ての局所特徴を一括で評価し最終的な確率を計算する．これは，$\prod_l^L$ で表される．各局所特徴に最も類似した部位を選択し $(\arg\max_r)$，その部位に対する確率を計算し，全局所特徴の確率の積として最終的な確率が計算される．次に，オクルージョン (具体的には，必要な局所特徴の欠落) への対応について述べる．Fergus の星座モデルはオクルージョンへの明示的な対応を行っている．部位と局所特徴の割当てを計算する際に，一部の部位には局所特徴を割り当てず，オクルージョンにより欠落した部位を表現する．また，隠れる部位の組み合わせの各頻度もモデル化する．Fergus の星座モデルでは，これら明示的な対応により，オクルージョンを考慮した確率が計算される．本手法では，最終的な確率は，画像中に存在している局所特徴から一括で計算される．従って，オクルージョンにより一部の局所特徴が検出されなかった場合，最終的な確率は，単純に，その局所特徴を除外した確率として適切に計算される．また，頻出するオクルージョンのパターン

(1) モデルパラメータ $\theta_{k,r}$, π_k を初期化.

(2) **E step** :

$$q_{k,n} = \frac{\pi_k p(I_n|\theta_k)}{\sum_k^K \pi_k p(I_n|\theta_k)} \qquad \text{ここで，} p(I_n|\theta_k) = \prod_l^L G(\mathbf{x}_{n,l}|\theta_{k,\hat{r}_{k,n,l}}).$$

(3) **M step** :

$$\boldsymbol{\mu}_{k,r}^{new} = \frac{1}{Q_{k,r}} \sum_n^N \sum_{l:(\hat{r}_{k,n,l}=r)} q_{k,n}\mathbf{x}_{n,l}$$

$$\boldsymbol{\Sigma}_{k,r}^{new} = \frac{1}{Q_{k,r}} \sum_n^N \sum_{l:(\hat{r}_{k,n,l}=r)} q_{k,n}\left(\mathbf{x}_{n,l} - \boldsymbol{\mu}_{k,r}^{new}\right)\left(\mathbf{x}_{n,l} - \boldsymbol{\mu}_{k,r}^{new}\right)^t$$

$$\pi_k^{new} = \frac{N_k}{N}$$

$$\text{ここで, } Q_{k,r} = \sum_n^N \sum_{l:(\hat{r}_{k,n,l}=r)} q_{k,n} \quad , \quad N_k = \sum_n^N q_{k,n}.$$

(4) パラメータ変化が収束したら終了．そうでなければ (2) へ.

図 1　マルチモーダル星座モデルのモデルパラメータ推定アルゴリズム

は，マルチモーダル化により，物体の一つの見た目として学習される．これは，隠れる部位の組み合わせの各頻度のモデル化と対応する．本手法では，この様に，暗黙的な対応により，オクルージョンを考慮した確率が計算される．最後に，オクルージョンとは逆に不要な局所特徴が含まれている場合を考える．Fergus の星座モデルでは，各部位に対して対応する局所特徴の網羅的な検査において，不要な局所特徴を含む割当てにおける確率はとても小さな値となる．しかし，最終的な確率は全ての割当てにおける確率の和で計算されるため，不要な局所特徴の影響は殆ど受けない．対して本手法では，全局所特徴の確率の積で計算されるため，不要な局所特徴は最終的な確率の値を下げる．しかし，それぞれの分類候補カテゴリにおいても同様に確率の値は下がり，従って，確率の値の比較として行われる分類処理において，分類結果は，不要な局所特徴の影響を受けない.

ここで，学習に必要な計算時間について，Fergus の星座モデルと比較する. [3] によると，Fergus の星座モデルの具体的な計算時間は，$R = 6$〜7，$L = 20$〜30，学習画像 400 枚の場合，一つのモデルの学習に 24〜36 時間かかるとのことである．しかし，上述した高速化の工夫を行った提案モデルでは，R, L, 学習画像枚数を同じ条件にし，$K = 1$(ユニモーダル) として学習を行った場合，数秒から十数秒ほどで学習が終了した．また，マルチモーダル化した場合（$K \geqq 2$）でも数十秒ほどで学習が終了した．この事から，本研究で用いた高速化の工夫はとても有効であったことが分かる.

2.4　モデルパラメータの推定

モデルパラメータの推定は EM アルゴリズム [12] で行う．マルチモーダル星座モデルにおけるモデルパラメータ推定アルゴリズムを図 1 に示す．なお，N は学習画像枚数を，n は画像インデックスを，$\hat{r}_{k,n,l}$ はインデックスが n の学習画像に対する $\hat{r}_{k,l}$ を表す.

混合ガウス分布に対する一般的な EM アルゴリズムと異なる点は，$\boldsymbol{\mu}, \boldsymbol{\Sigma}$ を更新するデータが，学習画像（混合ガウス分布における学習データ）単位ではなく，学習画像から得られた局所特徴単位であるという点である．各学習画像 n は，各構成要素 k への帰属確率 $q_{k,n}$ が計算され，局所特徴はその値に基づき $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ の更新に関与する．また，各画像 n の各局諸特徴 l は，構成要素 k での対応する部位 $\hat{r}_{k,n,l}$ の $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ の更新にのみ関与する.

3. 分類処理

分類処理は，$\hat{c}$ を分類結果のカテゴリ，c を分類候補のカテゴリとすると，

$$\hat{c} = \arg\max_c p_m(I|\Theta_c)p(c)$$

と記述できる．$p(c)$ は，カテゴリ c に関する事前確率であるが，これは各カテゴリの学習画像枚数の，全候補カテゴリに対する比率となる.

星座モデルは生成モデルであるため，新しいカテゴリの追加や分類候補カテゴリの変更は容易である．学習処理は，新しくカテゴリを追加する際に，それぞれのカテゴリとは独立に，一度のみ行えばよい．また，既に学習済みのカテゴリにおける分類候補カテゴリの変更は，分類処理に用いるモデルを差し替えるだけでよい．これに対して，判別モデルでは，学習に全ての候補カテゴリのデータを同時に用いて，一つの識別器（決定境界）を構築するため，候補カテゴリの追加や変更には必ず再学習が必要となる．この再学習は，候補カテゴリの追加や変更のたびに学習処理が発生するという欠点だけではなく，再学習を行うために，全候補カテゴリの学習データを保持しておかなくてはいけないという欠点も併せ持つ.

図 2　Caltech データセットの対象画像例

表 1　Caltech データセットの内訳

カテゴリ名	物体領域数
Airplanes	1074
Cars Rear	1155
Faces	450
Motorbikes	826

図 3　Pascal データセットの対象画像例

表 2　Pascal データセットの内訳

カテゴリ名	物体領域数
Bicycle	649
Bus	469
Car	1708
Cat	858
Cow	628
Dog	845
Horse	650
Motorbike	549
Person	2309
Sheep	843

4. 実　　験

星座モデルにおけるマルチモーダル化の有効性の評価のため，マルチモーダル星座モデル（Multi-CM）とユニモーダル星座モデル（Uni-CM）を比較する．Uni-CM は，$K = 1$ とした提案手法とする．

また，提案手法の性能を BoF を用いた二つの関連手法と比較する．LDA+BoF と，SVM+BoF である．それぞれ，文章解析手法の一つである LDA を用いた手法と，分類器の一つである SVM を BoF に適用した手法である．なお，Multi-CM，Uni-CM，LDA+BoF は生成モデルであり，SVM+BoF は判別モデルである．

次に，提案モデルにおける二つのハイパーパラメータであるモデルの要素数 K と部位数 R の変化に対する正答率の変化について考察を行う．さらに，1. 章で星座モデルの利点として挙げた (b) 連続値表現のため BoF よりも記述精度が高い，(c)BoF では無視した位置とスケール情報を適切に利用可能，を定量的に示す．

4.1 実験条件

実験で用いるタスクを変更する理由として，1 章で述べた事以外に，異なるタスクが設定されているデータセット間での結果の比較を可能にすることがある．本稿では，難易度の異なる二つのデータセットに対し実験を行い，本手法の評価を行う．実験の準備として，まず，対象とするデータセットに含まれる物体領域の情報を用いて画像から物体領域を切り出す．切り出した物体領域の画像を対象画像とする．これら対象画像を適切なカテゴリに分類することを本稿で行うタスクとする．この切り出し処理には，タスク変更の理由以外に，背景を排除することでデータセット間での正答率の違いの要因を見た目の多様性の違いのみにし，結果の見通しを良くするためという理由がある．なお，この切り出し処理は，本手法の必要要件ではない．

本稿で用いるデータセットは，Caltech Database [3]（以降，Caltech）と，一般物体認識のコンテストである PASCAL Visual Object Classes Challenge 2006 [13] で使用されたデータセット（以降，Pascal）である．Caltech には 4 カテゴリの画像が含まれる．表 1 に物体領域数の内訳を示し，図 2 に対象画像例を示す．対象画像中の物体は大まかに向きが揃えられているが，物体自体の見た目は大きく異なる．Pascal には，10 カテゴリの画像が含まれる．表 2 に物体領域数の内訳を示す．また，図 3 に対象画像例を示す．画像中の物体の向きはそれぞれ異なり，また物体自体の見た目も大きく異なる．さらにカテゴリによっては物体の姿勢も大きく異なる (例：Cat，Dog，Person など)．そのため，Caltech に比べ難易度が高いと考えられる．

局所特徴は，比較を行う全ての手法において同一の抽出済みデータを用いる．これは，抽出結果の違いによる正答率のばらつきを排除し，各手法の分類性能を厳密に比較するためである．対象画像の半分を学習に，残りをテストに用いる．学習とテストに用いる画像を変えて 10 回実験を行い，平均正答率で評価を行う．また，実験は対象とするデータセットごとに個別に行い，結果を比較する．

モデルの構成要素数 K は経験的に 5 とした．また，各構成要素中の部位数 R も同じく経験的に 21 とした．なおこれらハイパーパラメータの値を変化させたときの正答率の変化について 4.3 章および 4.4 章で考察する．

本稿では局所特徴として，検出には Kadir Brady saliency detector（以降，KB detector）[14] を，記述には DCT(Discrete Cosine Transform) を用いた．KB detector は局所特徴の位置と領域の大きさ（スケール）を出力する．その情報に基づき画像領域を切り出し，DCT で得られる直流を含まない最初の 20 個の係数を用いて，見た目の特徴ベクトルを表す．従って，特徴ベクトル $\mathbf{x}$ の次元数は，$\mathbf{A}$ が 20 次元，$\mathbf{X}$ が 2 次元，$\mathbf{S}$ が 1 次元であるので，合計 23 次元となる．

4.2 マルチモーダル化の効果と関連手法との比較

マルチモーダル化の効果を検証するため，Uni-CM と，Multi-CM の正答率を比較する．また，関連手法である LDA+BoF と SVM+BoF に対して比較を行う．なお，それぞれの関連手法にはハイパーパラメータとして，BoF における codeword の数 (k-means の k) と，LDA の想定トピック数が存在する．想定トピック数は，提案手法における構成要素数 K に対応する．なお，以下で示す比較結果では，これらのハイパーパラメータを変化させて複数の結果を取得し，その中で正答率が最も高かった結果を示す．

結果を表 3 に示す．Multi-CM の方が，Uni-CM より正答率が高い．これは，Caltech，Pascal 共に，カテゴリ中の物体には，大きく異なった複数の種類の見た目が存在したため（例：Caltech・Face：人物の違い，Pascal・Bicycle：自転車の向き），提案手法が有効であったことを示している．

また，LDA+BoF(生成モデル) と SVM+BoF(判別モデル) に対して，提案モデルはより高い正答率を示しいていることが分かる．この結果により，星座モデルを用いることで，生成モデル，判別モデル関わらず，BoF を用いた場合よりも高い正答

表 3　マルチモーダル化の効果と関連手法との比較．平均正答率 (%).

	LDA+BoF	SVM+BoF	Uni-CM	Multi-CM
Caltech	94.71	96.44	98.71	99.45
Pascal	29.62	27.90	37.02	38.77

率が得られることが示された．

また，Caltech の正答率と Pascal の正答率はどの手法を見ても同様に大きく異なっており，Pascal に含まれる物体の見た目の多様性が，Caltech のそれを大きく上回っていることが，正答率の違いから確認できる．

4.3　構成要素数 K の変化に対する正答率の変化

ここでは，提案手法のハイパーパラメータのひとつである要素数 K を変化させた場合の正答率の変化について考察を行う．K を 1 から 9 まで 2 づづ増加させ，各 K での正答率を比較する．K=1 は Uni-CM を，$K \geq 2$ は Multi-CM を意味する．部位数 R は 21 に固定する．

図 4 に結果を示す．なお，各グラフの縦軸のスケールは，データセットの難易度の違いから，それぞれのグラフで異なっているため注意してほしい．Caltech では，K の増加による正答率の改善は K=5 で飽和しており，Pascal では K=7 で飽和している．Pascal の方が飽和する K が大きい理由は，Pascal における物体の見た目の多様性が Caltech よりも大きいためである．しかしながら，要素数 K は一定値として設定することができる．$K \geq 2$ であれば，K の変化による正答率の変化はそれほど大きくないためである．従って，本稿では K=5 としている．

また実験結果は，$K \geq 2$ での正答率が K=1 の場合よりも高く，マルチモーダル化の有効性を示している．

4.4　部位数 R の変化に対する正答率の変化

提案手法のもう一つのハイパーパラメータ，部位数 R を変化させた場合の正答率の変化について考察を行う．R を 3 から 21 まで，3 ずつ増やしていき，各 R での正答率を調べる．構成要素数 K は 5 に固定する．正答率は，Uni-CM と Multi-CM，両方に対して求める．

結果を図 5 に示す．結果は，R の増加により正答率が向上し，Caltech では，$R = 9$ ごろ，Pascal では，$R = 21$ ごろ正答率の向上の飽和が見られた．Fergus の星座モデルでは，$R = 6$〜7 程度が現実的な計算時間で学習が終了する目安であったが，本手法では，処理の高速化により，正答率の向上が飽和するまで現実的な計算時間内で R の値を増加させることが可能となった．従って本稿で行った高速化の工夫は，マルチモーダル化の実現だけでなく，性能向上にも寄与したといえる．

また，いずれの R に対しても，Multi-CM の方が Uni-CM よりも正答率が高く，この結果からもマルチモーダル化の有効性を確認できた．

4.5　連続値表現と位置スケール情報の効果の検証

1. 章で述べた星座モデルの利点である (b) 連続値表現のため BoF よりも記述精度が高い，(c)BoF では無視した位置とスケール情報を適切に利用可能，を定量的に評価する．

まず，(b) について検証する．BoF と星座モデルの比較は，確率分布関数による連続値表現と，ヒストグラムによる離散表現の違いのみが残るように，できるだけそれ以外の条件を等しくして行う．そのために，星座モデルと同じ生成モデルである LDA+BoF と，LDA+BoF では用いない位置とスケール情報を使用しない Multi-CM（以降，Multi-CM no-X,S) を比較する．次に，(c) を検証するため，Multi-CM no-X,S と，通常の Multi-CM を比較する．

表 4 に，各 3 種類の手法の正答率を示す．結果は，LDA+BoF よりも Multi-CM no-X,S の方が正答率が高く，連続値表現の優位性を示している．また，Multi-CM no-X,S よりも，Multi-CM の方が正答率が高く，位置とスケール情報を星座モデルが適切に利用できていることを示している．

5. ま と め

一般物体認識のためのマルチモーダル星座モデルを提案した．提案したマルチモーダル星座モデルは，カテゴリ中の物体の見た目が複数の種類に分かれる場合でも高い精度でカテゴリを記述できる．しかしながら，マルチモーダル星座モデルの構造は Fergus の星座モデルよりもシンプルであり，実装が容易である．また，処理の高速化を行い，学習や識別が，Fergus の星座モデルよりも短時間で終了するようになった．

1. 章や 3. 章で述べたように，星座モデルは生成モデルであり，SVM 等の判別モデルによる手法に対する明らかな優位性を持つ．また，星座モデルは連続値表現であるため，離散表現である BoF よりも記述精度が高く，また，位置スケール情報を適切に利用可能である．

今後行っていく事として，高速化に関する正確な評価がある．今回は文章でのみ，本稿で行った高速化の工夫の妥当性を説明したが，今後は，分類性能への影響についてより詳しく調べていく予定である．

文　献

[1] 柳井啓司，“一般物体認識の現状と今後，” 情処学論，vol.48, no.SIG 16(CVIM 19)，pp.1-24，Nov. 2007.

[2] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” Proc. ECCV International Workshop on Statistical Learning in Computer Vision, pp.1-22, 2004.

[3] R. Fergus, P. Perona, and A. Zisserman, “Object class recognition by unsupervised scale-invariant learning,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.2, pp.264-271, 2003.

[4] K. Grauman, and T. Darrell, “The pyramid match kernel: discriminative classification with sets of image features,” Proc. IEEE Int. Conf. on Computer Vision, vol.2, pp.1458-1465, 2005.

[5] M. Varma, and D. Ray, “Learning the discriminative power-invariance trade-off,” Proc. IEEE Int. Conf. on Computer Vision, 2007.

[6] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local

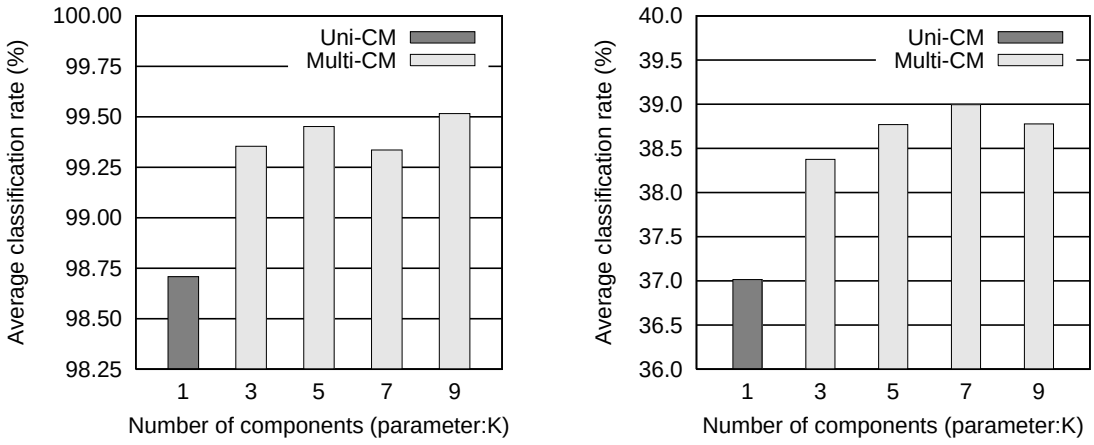

図 4　構成要素数 K の変化に対する正答率の変化

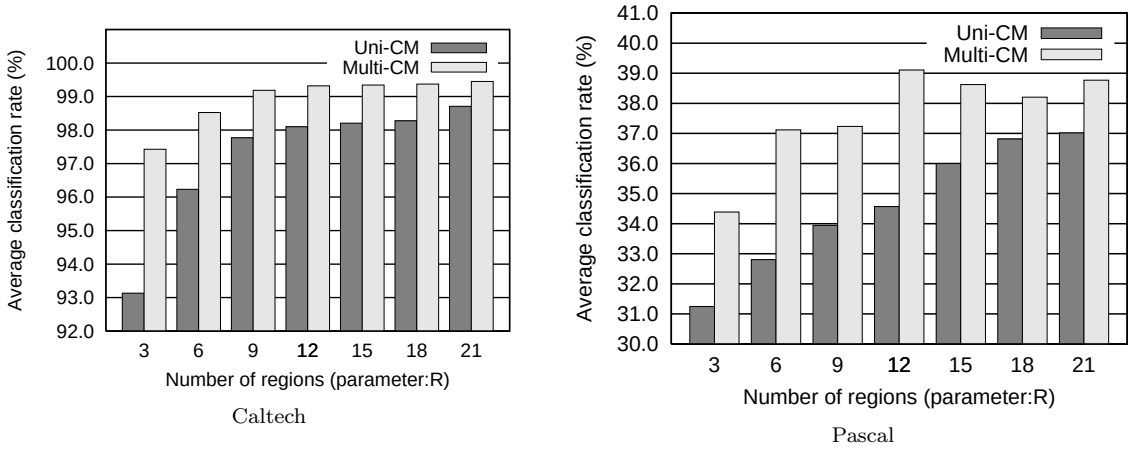

図 5　部位数 R の変化に対する正答率の変化

表 4　位置スケール情報と連続値表現の効果の検証．平均正答率 (%).

	LDA+BoF	Multi-CM no-X,S	Multi-CM
Caltech	94.71	96.47	99.45
Pascal	29.62	33.47	38.77

features and kernels for classification of texture and object categories: A comprehensive study," Int. J. of Computer Vision, no.2, pp.213-238, 2007.

[7] A. Bosch, A. Zisserman, and X. Munoz, "Scene classification via pLSA," Proc. European Conf. on Computer Vision, vol.4, pp.517-530, 2006.

[8] L. Fei-Fei, and A.P. Perona, "A bayesian hierarchical model for learning natural scene categories," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.2, pp.524-531, 2005.

[9] G. Wang, Y. Zhang, and L. Fei-Fei, "Using dependent regions for object categorization in a generative framework," Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.2, pp.1597-1604, 2006.

[10] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[11] X. Ma, and W.E.L. Grimson, "Edge-based rich representation for vehicle classification," Proc. IEEE Int. Conf. on Computer Vision, vol.2, pp.1185-1192, 2005.

[12] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Statistical Society, Series B, vol.39, no.1, pp.1-38, 1977.

[13] M. Everingham, A. Zisserman, C.K.I. Williams, and L. Van Gool, "The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results," http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf.

[14] T. Kadir, and M. Brady, "Saliency, scale and image description," Int. J. of Computer Vision, vol.45, no.2, pp.83-105, November 2001.