一般物体認識のためのマルチモーダル星座モデル*

神谷　保徳†　　高橋　友和††　　井手　一郎†　　村瀬　洋†

A Multimodal Constellation Model for Generic Object Recognition*

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

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ー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．この分野で現在主流になりつつある手法として Bag of Features (BoF) を用いた手法と，星座モデルがある．星座モデルにはいくつかの利点があるが，欠点として，本質的にユニモーダル（単峰性）なモデルであるため，カテゴリー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精度が低くなるということがある．本論文では，この欠点を解決する“マルチモーダル星座モデル”を提案する．マルチモーダル化により物体のそれぞれの種類の見た目を個別に記述できるようになり，記述精度が向上する．実験ではユニモーダルなモデル及び BoF を用いた手法に比べ識別性能が高いことを確認した．

キーワード　星座モデル，マルチモーダル化，一般物体認識，EM アルゴリズム

1. まえがき

我々人間は，車，椅子，焼そばなどの，物体が属するカテゴリー（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よ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認識を行うことは，現在の技術では困難である．カテゴリー中には，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 [4]〜[6] や，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 [7]〜[9] がある．

一方星座モデルは，対象カテゴリー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ー内の物体に共通する部位の見た目とその位置関係，部位領域の大きさを記述する．星座モデルには，以下の三つの利点がある．

(a)　処理対象カテゴリー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識別的アプローチ（識別モデル + 識別関数）[10] に分類される．この利点は，星座モデルが生成モデルであることに起因する．生成モデルは，処理対象カテゴリーをそれぞれ独立にモデル化するため，カテゴリーの追加や変更に必要な作業は，そのカテゴリーに対してのみ行えばよい．これに対して，識別的アプローチはすべての処理対象カテゴリーを区別する決定境界の最適化として学習されるため，全カテゴリーにわたった作業

† 名古屋大学大学院情報科学研究科，名古屋市
Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464–8601 Japan
†† 岐阜聖徳学園大学経済情報学部，岐阜市
Faculty of Economics and Information, Gifu Shotoku Gakuen University, 1–38 Nakauzura, Gifu-shi, 500–8288 Japan
* 本論文は第 12 回画像の認識・理解シンポジウム推薦論文である．

が必要となる．

(b)　連続値表現であるため記述精度が高い：BoF によるカテゴリーの記述は離散表現である．特徴空間中の局所特徴の分布をベクトル量子化（codeword 化）して離散値で扱い，それぞれの codeword に対する局所特徴の個数からなるヒストグラム（離散値）で記述する．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り，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．これに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，カテゴリー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，本質的にユニモーダル（単峰性）なモデルであり，そのため，カテゴリー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，記述精度が低いという欠点がある．

本論文では，この欠点を改善するため “マルチモーダル星座モデル” を提案する．まず，この欠点に対して，モデルをマルチモーダル（多峰性）に拡張する．マルチモーダルなモデルに拡張することで，複数の種類の見た目を個別に記述可能にし，カテゴリー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，局所特徴を用いて記述を行うモデルに対して行ったことに等しい．このマルチモーダル化により記述精度は向上するが，星座モデルはモデルパラメータの推定に時間がかかり，マルチモーダル化によりモデルパラメータが増加することで，現実的な時間内での学習終了が非常に困難になる．そこで，処理の高速化も併せて行う．

一般物体認識において基本的なタスクである，単一の物体を写した画像の適切なカテゴリーへの分類処理を，本論文では対象とする．物体画像の適切なカテゴリーへの分類処理は，一般物体認識における本質的な課題（見た目の多様性に基づく困難さ）を含む非常に基本的なタスクである．

以降，**2.** でマルチモーダル星座モデルについて述べる．**3.** 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.** で実験について述べ，**5.** で本論文をまとめる．

1.1 関 連 研 究

星座モデルと呼ばれるモデルは，本論文で対象とする Fergus のモデル以前に，Weber によっても提案されている [11]．星座モデルのマルチモーダル化は，Weber の星座モデルで行われてはいる [12] が，Weber のモデルのモデル構造は Fergus のモデルとは全くの別物であり，三つの欠点，つまり，局所特徴の扱い方が BoF に近く利点 (b) を保持していないこと，部位領域の大きさを扱うことができないこと，見た目と位置関係の学習が同時にはできずそれぞれの学習に依存関係があること，が存在する．

また，Fergus の星座モデルを改良したモデルも存在する [13]．このモデルでは，複数の種類の局所特徴を利用可能にしたり，位置関係の記述の改良を行っているが，マルチモーダル化は行われておらず，そこで本論文では Fergus の星座モデルのマルチモーダル化を行う．

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]

カテゴリーに共通する物体の複数の部位に着目し記述を行う．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のとおりである．

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

ここで，I は入力画像，Θ はモデルパラメータであり，$\Theta = \{\theta_A, \theta_X, \theta_S, \theta_{other}\}$ 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 から得られたすべての局所特徴を，モデルが表現する各部位に割り当てる割り当て方の一つを表現するベクトルであり，H は，割り当て方のすべての組合せの集合である．$\sum_{\mathbf{h}\in H}$ 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当の組合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．詳細は文献 [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当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，対象カテゴリーを表現する部分 $(p(A,X,S,\mathbf{h}|\Theta))$ 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

本論文で提案するマルチモーダル星座モデルを以下のように定義する．

$$
\begin{aligned}
p_m(I|\Theta) &= \sum_k^K \left\{ \prod_l^L G(\mathbf{x}_l|\theta_{k,\hat{r}_{k,l}}) \right\} \cdot \pi_k \\
&= \sum_k^K \left\{ \prod_l^L G(\mathbf{A}_l|\theta^{(A)}_{k,\hat{r}_{k,l}}) \right. \\
&\qquad \left. \cdot\, G(\mathbf{X}_l|\theta^{(X)}_{k,\hat{r}_{k,l}}) G(\mathbf{S}_l|\theta^{(S)}_{k,\hat{r}_{k,l}}) \right\} \cdot \pi_k \\
\hat{r}_{k,l} &= \arg\max_r G(\mathbf{x}_l|\theta_{k,r}).
\end{aligned}
$$

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，$G(\,)$ は平均 $\boldsymbol{\mu}$，分散 $\boldsymbol{\Sigma}$ のガウス分布を表す．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 の存在確率であり，$0\le\pi_k\le 1$ と $\sum_k^K \pi_k=1$ を満たす．$\hat{r}_{k,l}$ 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，式中には現れないが，このほかにハイパパラメータとして部位数 R がある．

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量は大きく，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．このことは，星座モデルのマルチモーダル化の実現を困難にする．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困難になるためである．そこで，処理の高速化が必要となる．ここでは本手法において行った，2 点の高速化の工夫について述べる．

［1. 行列計算の簡略化］ モデル中のすべての共分散行列 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．Σ を $D\times D$ 行列とすると，計算量は，$O(D^3)$ 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，物体カテゴリー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，$\boldsymbol{\Sigma}$ の対角要素，$\boldsymbol{\Sigma}$ の非対角要素，の三つに分けることができる．平均と $\boldsymbol{\Sigma}$ の対角要素は分布ごとに異なる値をもつ．しかし $\boldsymbol{\Sigma}$ の非対角要素は，各特徴変数が独立に近いほど 0 に近い値となる．一般的に分類処理において特徴変数はできるだけそれぞれが独立なものを用いるため，分類処理に対する $\boldsymbol{\Sigma}$ の非対角要素の影響度は，平均や対角要素に比べ低い．

［2. $\Sigma_{h\in H}$ 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を網羅的に計算する部分 $\sum_{\mathbf{h}\in H}$ 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 [14] 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，まず Fergus の手法と本手法のモデル構造を計算手順とともに説明し，その後以下の 2 点について両手法を比較しつつ述べ，この高速化の工夫が最終的な結果に与える影響が小さいことを示す．

- オクルージョン（必要な局所特徴の欠落）への対応
- 不要な局所特徴が含まれる場合への対応

まず，両手法のモデル構造を計算手順とともに説明

(1) モデルパラメータ $\theta_{k,r}(=\{\boldsymbol{\mu}_{k,r}, \boldsymbol{\Sigma}_{k,r}\})$, 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
Fig. 1　Model parameter estimation algorithm for the Multimodal Constellation Model.

する．Fergus の星座モデルは，それぞれの部位に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，すべての局所特徴を一括して評価し最終的な確率を計算する．これは，$\prod_l^L$ で記述される．各局所特徴に最も類似した部位を選択し ($\arg\max_r$)，各局所特徴の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を計算する．

次に，オクルージョン（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，欠落する部位の組合せごとの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．これに対して本手法では，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して計算される．したがって，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，隠れる部位の組合せごとの頻度のモデル化と対応する．本手法では，このように，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，物体の記述には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 **h** における確率は小さな値となる．最終的な確率はすべての割当における確率の和で計算されるため，その結果，不要な局所特徴の影響はほとんど受けない．これに対して本手法では，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ーにおいてもこの局所特徴は記述に不要であるため，ほぼ同様に確率の値は下がり，したがって，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ほとんど受けない．

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 [15] 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，$\mathbf{x}_{n,l}$ は学習画像 n における局所特徴 l の特徴ベクトルを，$\hat{r}_{k,n,l}$ は学習画像 n における $\hat{r}_{k,l}$（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）を表す．

(1) の初期化処理における初期値について説明する．$\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$（対角行列，対角成分 σ^2 のみ）の初期値は，特徴量の値がとり得る範囲に基づいて初期化する．$\boldsymbol{\mu}$ は，特徴量の値の範囲を考慮した範囲内のランダムな値で初期化する．$\boldsymbol{\Sigma}$ は，特徴量の値の範囲を考慮した固定値で初期化する．また，π については，$\frac{1}{K}$ で初期化する．

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 ごとに，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，$\boldsymbol{\mu}, \boldsymbol{\Sigma}$ の更新には，$\hat{r}_{k,n,l}$ が r となる l のみ関与する（$\sum_{l:(\hat{r}_{k,n,l}=r)}$ の箇所）．

3. 分類処理

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

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

と記述できる．$p(c)$ は，カテゴリー c に関する事前確率であるが，これは各カテゴリーの学習画像枚数の，全候補カテゴリー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，新しいカテゴリーの追加や分類候補カテゴリーの変更は容易である．処理対象カテゴリーはそれぞれ個別にモデル化される．学習処理は，新しくカテゴリーを追加する際に，それぞれのカテゴリーとは独立に，一度のみ行えばよい．また，既に学習済みのカテゴリーにおける分類候補カテゴリーの変更は，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，識別的アプローチでは，学習にすべての候補カテゴリー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，候補カテゴリー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，候補カテゴリー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，全候補カテゴリー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．

4. 実　　験

星座モデルにおけるマルチモーダル化の有効性を評価するため，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 は識別的アプローチである．また，LDA はマルチモーダルなモデルである．

次に，提案モデルにおける二つのハイパ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の考察と，各構成要素で記述される物体の見た目に関する考察を行う．

また，本手法で行った高速化により，モデルの記述能力が低下しているという懸念を排除するため，本手法と Fergus の星座モデルの比較を行う．また，学習時間に関する考察も行う．

最後に，**1.** で星座モデルの利点として挙げた利点のうち，(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 Multi-CM においても保持していることを示す．

4.1 実験条件

本実験ではデータセットとして，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 [16] で使用されたデータセット（以降，Pascal）を用いた．本実験で用いるデータセットとして望ましいものは，カテゴリー内で見た目のバリエーションがはっきりと存在し，かつ，カテゴリー内のそれぞれの見た目に属する画像枚数を十分確保できる（= 各カテゴリーの画像枚数が十分多い）データセットであるが，これらデータセットは，この条件を満たしている．ただしこれらデータセットは，本論文で対象とするタスク（単一の物体を写した画像の適切なカテゴリーへの分類処理）を想定したものではないため，あらかじめ

表 1　Caltech [3] の物体領域数の内訳
Table 1　Number of object areas in Caltech [3].

カテゴリー名	物体領域数
Airplanes	1,074
Cars Rear	1,155
Faces	450
Motorbikes	826

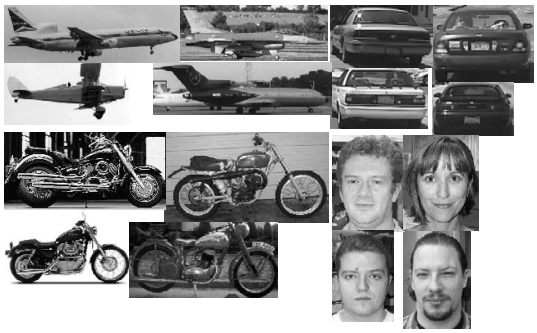

図 2　対象画像例：Caltech [3]
Fig. 2　Target images in Caltech [3].

これらデータセット内の画像を，対象タスクに合わせて修正した．各データセットに含まれる物体領域の情報を用いて，画像中の物体領域を切り出し，対象画像とした．これら対象画像について，適切なカテゴリーへの分類処理を行う(注1)．なお評価は，分類先カテゴリーの正答率で行う．

Caltech には 4 カテゴリーの画像が含まれる．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 カテゴリーの画像が含まれる．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．更にカテゴリー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，Pascal は Caltech に比べ難易度が高いと考えられる．

局所特徴は，比較を行うすべての手法において事前に抽出しておいた同一のデータを用いる．これは，正答率の比較において特徴量の影響を排除することで，各手法の分類性能をより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．

モデルの構成要素数 K は経験的に 5 とした．また，

表 2　Pascal [16] の物体領域数の内訳
Table 2　Number of object areas in Pascal [16].

カテゴリー名	物体領域数
Bicycle	649
Bus	469
Car	1,708
Cat	858
Cow	628
Dog	845
Horse	650
Motorbike	549
Person	2,309
Sheep	843

図 3　対象画像例：Pascal [16]
Fig. 3　Target images in Pascal [16].

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** 及び **4. 4** で考察する．

局所特徴の検出には Kadir Brady saliency detector（以降，KB detector）[17] 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を出力する．その情報に基づき画像領域（グレースケール）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した．したがって，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 BoF を用いた手法との比較

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，LDA+BoF と SVM+BoF との正答率の比較も行う．なお，これ

(注1)：本論文で対象とするタスクと同様のタスクを想定したデータセットとして，Caltech101，256 が存在するが，各カテゴリーの画像枚数が少なく本実験には適さなかった．

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 (%)
Table 3　Effectivity of multimodalization and comparison to related works, by average classification rates (%).

	LDA+BoF	SVM+BoF	Uni-CM	Multi-CM
Caltech	94.7	96.4	98.7	99.5
Pascal	29.6	27.9	37.0	38.8

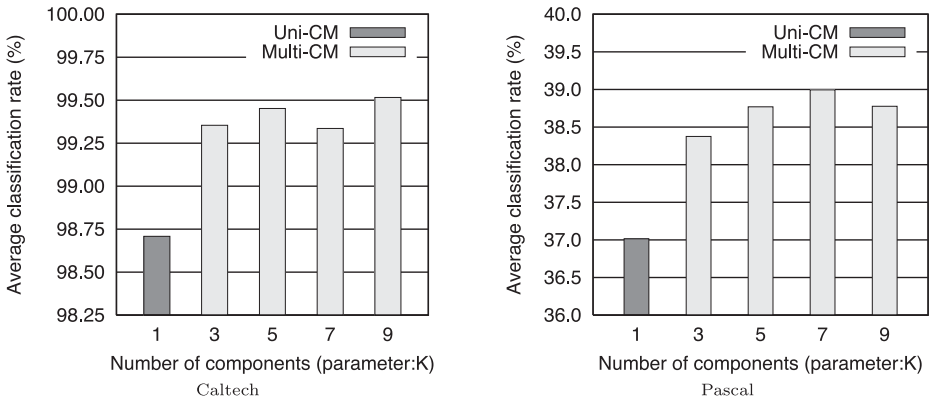

図 4　構成要素数 K の変化に対する平均正答率の変化
Fig. 4　Influence of K (number of components) on average classification rate.

らの手法にはハイパパラメータとして，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．LDA の想定トピック数は，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比較対象として示す．

結果を表 3 に示す．Caltech，Pascal ともに，Multi-CM の方が，Uni-CM より正答率が高い．これは，同一カテゴリー中の物体に存在する，大きく異なる複数の見た目（例：Caltech・Face：人物の違い，Pascal・Bicycle：自転車の向き）に対して，マルチモーダル化が有効に働いたことが要因と考えられる．**4.5** で示すが，マルチモーダル化により複数となった，モデル内の各構成要素によって，カテゴリー内の見た目をそれぞれ個別に記述できるようになり，それによりカテゴリーの記述精度が向上し，正答率の向上につながったと考えられる．

また，LDA+BoF（生成モデル）と SVM+BoF（識別的アプローチ）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，識別的アプローチにかかわらず，BoF を用いた手法よりも高い正答率が得られることが示された．

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，提案手法のハイパパラメータの一つである要素数 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 ずつ増加させ，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が必要である．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 $K = 7$ で飽和していると考えられる．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，$K \geq 2$ 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，本論文では $K = 5$ としている．

また $K \geq 2$ での正答率が $K = 1$ の場合よりも高いことは，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．Multi-CM の構成要素数 K は 5 に

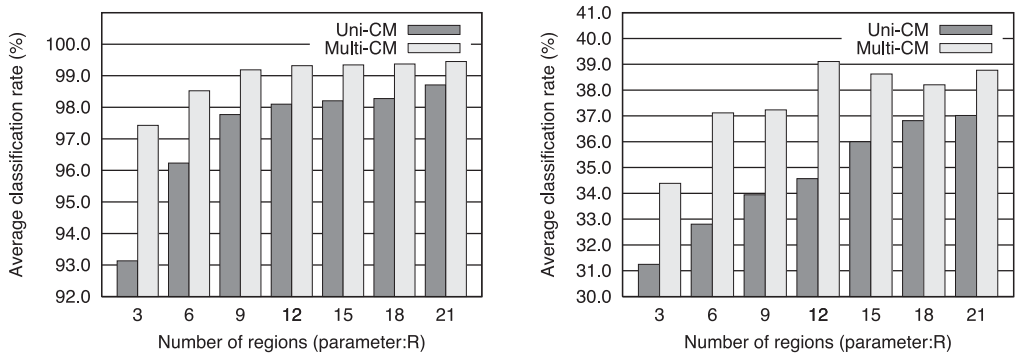

図 5　部位数 R の変化に対する平均正答率の変化
Fig. 5　Influence of R (number of regions) on average classification rate.

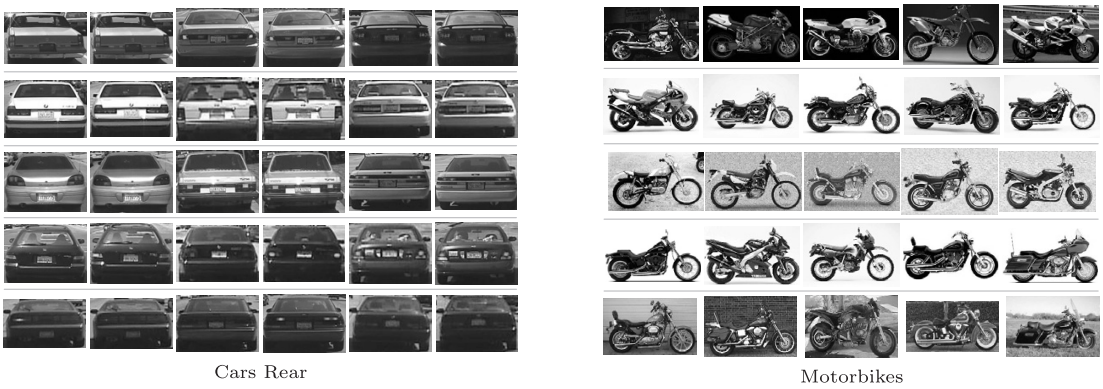

図 6　各構成要素に属する画像例 (Caltech)（各行が構成要素を表す）
Fig. 6　Example of groupings for each component of the model (Caltech). Each row shows each component.

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．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．モデルを構成する部位には，カテゴリー内のどのような見た目でも共通して使用される主要な部位とそうでない付属的な部位が存在する．見た目ごとのグループを記述する構成要素内では，付属的な部位の入換えで見た目の多様性が表現される．また，付属的な部位の欠落によりオクルージョンが表現される．Pascal は見た目の多様性が大きく，付属的な部位がより多く必要となるため，Caltech と比較して正答率が飽和する R が大きくなると考えられる．

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　各構成要素で記述される物体の見た目に関する考察

学習の結果，モデルの構成要素として記述された物体の見た目について考察する．どのような見た目がそれぞれの構成要素で記述されたのかを分かりやすくするため，構成要素数 K を少し大きめの 10 としてモデルを学習し，その結果を用いて考察を行う．各カテゴリーのマルチモーダル星座モデルを，学習したカテゴリーと同じカテゴリーの検証用画像に適用し，モデル出力値に対する各構成要素の寄与率 $\left\{\prod_l^L G(\mathbf{x}_l|\theta_{k,\hat{r}_{k,l}})\right\}\cdot\pi_k$ を計算する．寄与率が最も大きな構成要素を，その検証用画像が属する構成要素とする．

構築された 10 個の構成要素のうち主要な構成要素五つについて，各構成要素に属する画像例を図 6 と図 7 に示す．Caltech の Cars Rear では，主に車種によってグループが構成されている．Motorbikes では逆に車種ではなく，主に背景の違いによってグルー

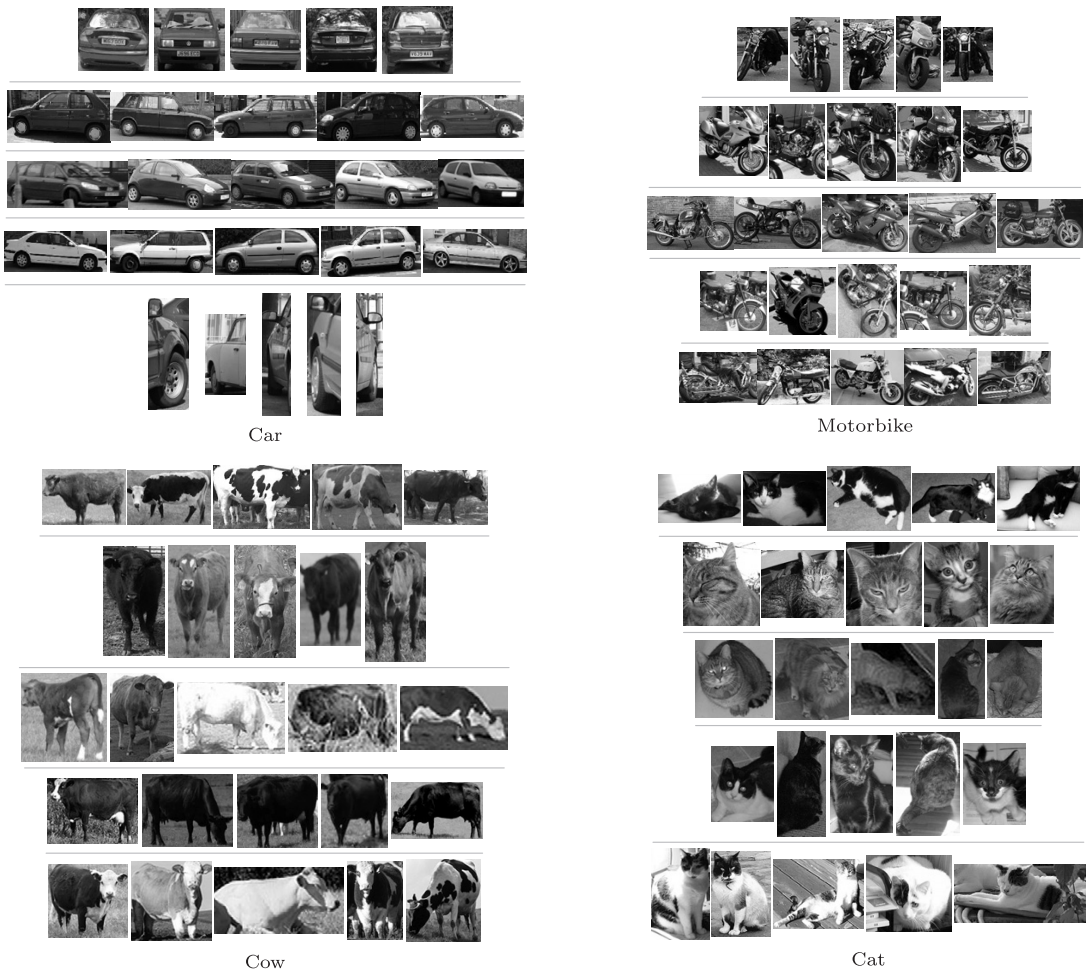

図 7　各構成要素に属する画像例 (Pascal)（各行が構成要素を表す）
Fig. 7　Example of groupings for each component of the model (Pascal). Each row shows each component.

プが構成されている．これは，車種の違いによる見た目の違いよりも，背景の違いによる見た目の違いの方が大きかったためである．Pascal の Car では主に車の向きと車両の色の明るさによってグループが構成されている．色の明るさが影響を与える理由は，局所特徴の記述に輝度値の DCT を用いているためである．Motorbike も向きがグループ構成に大きく寄与している．Cow や Cat は見た目のバリエーションが特に広く，グループが構成されにくいが，おおよそ体の向きや模様でグループが構成されていることが分かる．

4. 6　Fergus の星座モデルとの比較

本手法で行った高速化により，モデルの記述能力が低下しているという懸念を排除するため，本手法と Fergus の星座モデルの比較を行う．Fergus の星座モデルは計算量が膨大であるため，使用する局所特徴の画像 1 枚当りの個数 (L) を 20 に制限する．また，部位数 (R) についても，3 に設定する．本手法についても，同じ条件で実験を行う．なお，高速化による記述能力の低下に関する比較実験であるため，本手法は，マルチモーダル化しないモデル ($K = 1$) とする．前述の実験同様，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った．

実験結果を表 4 に示す．Caltech，Pascal ともに，本手法の方が，Fergus の星座モデルに比べ正答率が高かった．この結果から，高速化による記述能力への影響は少ないと考える．またこの結果から，実装の違い

表 4　Fergus の星座モデルとの比較．平均正答率 (%)．$L = 20$，$R = 3$, ($K = 1$，提案手法のみ)
Table 4　Comparison with Fergus's constellation model, by average classification rate (%). $L = 20$, $R = 3$, ($K = 1$, proposed model only).

	Proposed model	Fergus's model
Caltech	93.0	71.1
Pascal	31.3	19.5

表 5　連続値表現と位置スケール情報の効果の検証．平均正答率 (%)
Table 5　Validation of effectivity of continuous value expression and position-scale information, by average classification rate (%).

	LDA+BoF	Multi-CM no-X,S	Multi-CM
Caltech	94.7	96.5	99.5
Pascal	29.6	33.5	38.8

による性能の違いを想定しても少なくとも Fergus の星座モデルと同程度の識別性能を，本手法を用いて実現できることが示された．

4.7　学習時間の考察

ここで，学習に必要な計算時間について考察する．文献 [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り，マルチモーダル化を考えた場合，高速化の工夫なしでは現実的な時間内での学習終了は困難であることが分かる．

高速化の工夫を行った提案モデル ($K \geq 2$) では，R, L，学習画像枚数を同じ条件にし，数十秒ほどで学習が終了しており，現実的な時間内での学習終了が実現できていることが分かる．

また参考として，**4.6** での学習時間 ($R = 3$，$L = 20$，$K = 1$) を示す．提案手法では大まかに，1 モデル当り 1 秒程度，Fergus のモデルでは，5 分程度であった．

4.8　連続値表現と位置スケール情報の効果の検証

1. 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，を提案したマルチモーダル星座モデルについて評価する．

まず，利点 (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，提案手法と同じ生成モデルでマルチモーダルである LDA+BoF と，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，利点 (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 5 に，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マルチモーダル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. む す び

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，カテゴリー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ーを記述できる．実験では，マルチモーダル化の有効性と，BoF を用いた手法よりも高い識別性能を提案モデルがもっていることを示した．また，提案モデルを用いることで，高速な計算処理を実現しつつ，Fergus の星座モデルと少なくとも同程度の識別性能を実現できることを示した．

今後は，対象物体ごとに異なると考えられる見た目のバリエーションと構成要素数 K や部位数 R といったハイパパラメータとの関連性についてのより詳細な検証と，物体検出への本手法の応用を行っていく予定である．

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, 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 features and kernels for classification of texture and object categories: A comprehensive study,” Int. J. Comput. Vis., vol.73, 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] M. Weber, M. Welling, and P. Perona, “Unsupervised learning of models for recognition,” Proc. European Conf. on Computer Vision, vol.1, pp.18–32, 2000.
[12] M. Weber, M. Welling, and P. Perona, “Towards automatic discovery of object categories,” Proc. European Conf. on Computer Vision, vol.2, pp.101–108, 2000.
[13] R. Fergus, P. Perona, and A. Zisserman, “A sparse object category model for efficient learning and exhaustive recognition,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.1, pp.380–387, 2005.
[14] 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.
[15] 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.
[16] 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.
[17] T. Kadir and M. Brady, “Saliency, scale and image description,” Int. J. Comput. Vis., vol.45, no.2, pp.83–105, 2001.

（平成 20 年 10 月 10 日受付，21 年 2 月 25 日再受付）

神谷　保徳　（学生員）

平 17 名大・工・電気電子情報卒．平 19 同大大学院工学研究科博士前期課程了．現在，同大学院情報科学研究科博士後期課程在学中．物体認識・コンピュータビジョンの研究に従事．平 21 MMM2009 最優秀論文賞．情報処理学会会員．

高橋　友和　（正員）

平 9 茨城大・工・情報卒．平 12 同大大学院理工学研究科博士前期課程了．平 15 同研究科博士後期課程了．博士（工学）．同年より 2 年間名古屋大学大学院情報科学研究科 COE 研究員．平 17 より 3 年間日本学術振興会特別研究員．平 20 より岐阜聖徳学園大学経済情報学部准教授，現在に至る．画像認識の基礎研究並びにその応用に興味をもつ．画像電子学会会員．

井手　一郎　（正員）

平 6 東大・工・電子卒．平 8 同大大学院工学系研究科情報工学専攻修士課程了．平 12 同研究科電気工学専攻博士課程了．博士（工学）．同年国立情報学研究所助手．平 16 名古屋大学大学院情報科学研究科助教授，情報・システム研究機構国立情報学研究所客員助教授兼任．平 19 より准教授．この間，平 14〜16 総合研究大学院大学数物科学研究科助手併任，平 17〜19 フランス情報学・統計システム研究所（IRISA）招聘教授．パターン認識技術の実応用や映像メディア処理全般に興味をもっている．情報処理学会，人工知能学会，画像情報学フォーラム，IEEE Computer Society，ACM 各会員．

村瀬　　洋　（正員：フェロー）

昭 53 名大・工・電気卒．昭 55 同大大学院修士課程了．同年日本電信電話公社（現 NTT）入社．平 4 から 1 年間米国コロンビア大客員研究員．平 15 から名古屋大学大学院情報科学研究科教授，現在に至る．文字・図形認識，コンピュータビジョン，マルチメディア認識の研究に従事．工博．昭 60 本会学術奨励賞，平 6 IEEE-CVPR 最優秀論文賞，平 7 情報処理学会山下記念研究賞，平 8 IEEE-ICRA 最優秀ビデオ賞，平 13 高柳記念奨励賞，平 13 本会ソサイエティ論文賞，平 14 本会業績賞，平 15 文部科学大臣賞，平 16 IEEE Trans. MM 論文賞，ほか受賞．IEEE フェロー，情報処理学会会員．