

一般物体認識のためのタイプの異なる局所特徴の統合利用

神谷 保徳[†] 矢野 良和^{††} 大熊 繁^{†††} 高橋 友和^{††††}
井手 一郎[†] 村瀬 洋[†]

Combining Different Types of Local Features for Generic Object Recognition

Yasunori KAMIYA[†], Yoshikazu YANO^{††}, Shigeru OKUMA^{†††},
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では、画像中の局所的な領域を記述する局所特徴を用いた手法が一般的であり、これまでに様々な局所特徴が提案されている。本研究では、着目する領域の違いによりこれまで提案されてきた局所特徴を3種類に分類し、一般物体認識における認識性能向上のため、3種類の局所特徴を統合して用いる認識手法を提案する。我々は局所特徴を、(1) 特徴的なテクスチャを含む領域を扱う局所特徴、(2) 色や模様が一様な領域を扱う局所特徴、(3) 部分的なエッジ線を扱う局所特徴の3種類に分類する。実験では3種類の局所特徴を用いることの有効性を示す。キーワード 局所特徴、特徴的テクスチャ、色模様一様領域、エッジ情報、一般物体認識

1. ま え が き

ITS, セキュリティ, ロボティクスなど様々な分野で、人や車などの物体を画像から認識する物体認識の研究開発が行われている。我々人間は、物体が属するカテゴリー(例: 車, コンピュータ, アイス, などの一般名称)を、撮影環境が制約されない条件下においても容易に認識することができる。しかし現在の物体認識技術でこれを行うことは容易ではない。認識対象となるカテゴリーに属する物体のアピランスの変化が極めて大きく、特徴抽出, モデルの構築, 学習データセットの構築が困難なことが理由として挙げられる。このような物体認識は特に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考えられて

いる[9]。

現在一般物体認識の分野では、画像中の物体の複数の部分領域から抽出される局所特徴を用いる手法が目ざされ広く研究されている。これまで一般物体認識に使用されてきた局所特徴は、局所特徴が記述する領域の違いに基づく以下3種類に大別できる。

- タイプ1 特徴的なテクスチャに着目
- タイプ2 画像領域の一様性に着目
- タイプ3 エッジ情報に着目

また、局所特徴を画像から抽出する手法は、抽出する部分を特定する“検出”と、検出された領域の“アピランスの記述”に分かれ、各タイプ別にそれぞれ異なる手法が用いられている。表1は、各タイプの手法の一例を、局所特徴の検出とアピランスの記述別に分類したものである。なお本論文では、検出とアピランスの記述を合わせた処理を、局所特徴の抽出、と呼ぶことにする。タイプ1は、テクスチャが特徴的な(輝度変化が急で、一様でない)領域に着目する局所特徴である。DoGやHarris-Affine detectorなどの手法でこのような領域を検出し、SIFT記述子やDCTなどでその領域のテクスチャパターンを記述する。タイプ1の局所特徴を使用している研究を挙げる。[10]では、検出にHarris-Affine detectorが、記述にSIFT記述子が使用されている。[11],[12]では、検出にKadir

[†]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

^{††}愛知工業大学電気学科, 豊田市
Electronics Course, Aichi Institute of Technology, 1247
Yachigusa, Yakusa-cho, Toyota-shi, 470-0392 Japan

^{†††}名古屋大学大学院工学研究科, 名古屋市
Graduate School of Engineering, Nagoya University, Furo-
cho, Chikusa-ku, Nagoya-shi, 464-8603 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

表 1 局所特徴手法の一例の分類
Table 1 Grouped local feature methods.

局所特徴のタイプ	検出	アピランスの記述
タイプ 1: 特徴的なテクスチャに着目	DoG [1], Harris-Affine detector [2], Kadilar Brady saliency detector [3]	SIFT 記述子 [1], DCT, PCA
タイプ 2: 画像領域の一樣性に着目	領域分割手法 (mean-shift, blobworld [4], JSEG [5]), MSER [6], [7]	色 (平均値, ヒストグラム), texton [8]
タイプ 3: エッジ情報に着目	エッジフィルタ (Canny 等) + セグメンテーション	エッジ線の角度, 長さ, エッジ画像

Brady saliency detector (以降, KB detector) が, アピランスの記述に DCT や PCA が使用されている. タイプ 2 は, タイプ 1 とは対照的に, 輝度変化やテクスチャの変化がほとんどないか, 緩やかである領域を記述対象とする局所特徴である. 色や輝度値, 若しくはテクスチャが一樣な領域を, mean-shift などの領域分割手法や MSER で検出し, 得られた領域内に均一に広がる画像特徴を, 領域内の平均色やヒストグラム, texton などを用いて記述する. mean-sift 及び texton は [13] で利用されている. MSER は [14] などで, また, 輝度値が一樣な楕円領域を検出する手法である [7] は [15] で使用されている. タイプ 3 は, 色変化が急な境界を, 連続した線形状として着目する局所特徴である. Canny edge filter などのエッジ検出フィルタで得られたエッジ線を適当な手法で複数のセグメントに分割することで, タイプ 3 の局所特徴が検出される. 記述には, エッジ線の角度や長さ, エッジ線全体を画像として保持する方法が用いられる. [16] ではエッジ線の角度と長さが, [17] ではエッジ線の角度が, [18] ではエッジ画像が用いられている.

各タイプが検出する領域の違いの視覚的な理解を促すため, 図 1 に各タイプの局所特徴の検出領域のイメージ図を示す. 上の図の入力画像パターンから各タイプの局所特徴を検出した結果が下の図となる. タイプ 1 と 2 の検出領域は円領域で示し, タイプ 3 は黒線で示している.

一般物体認識は, 様々なカテゴリーを対象とした物体認識であり, また, 対象となるカテゴリーに含まれる物体の見目は幅広いため, 異なる性質をもった局所特徴を組み合わせることで物体を記述することが望まれる. これまで, いくつかの論文で, 異なるタイプの局所特徴を組み合わせる手法が提案されている. [19] や [20], [21] ではタイプ 1 とタイプ 3 を用いている. また, [14] ではタイプ 1 とタイプ 2 を用いている. 幅広い物体の記述を行うためにはすべてのタイプの局所特徴を用いることが望ましいが, まだそのような手法は

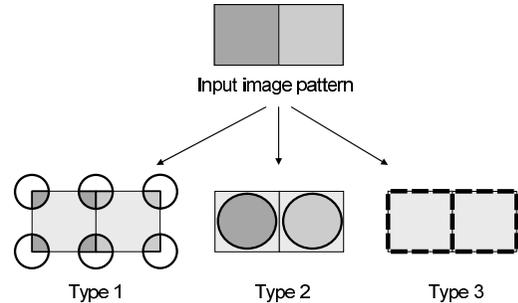


図 1 各局所特徴タイプによる検出領域のイメージ図
Fig. 1 Detected area imageries by each local feature type.

提案されていない.

また, 異なるタイプの組合せ方として, [19] では, 識別器を各タイプ別に独立に学習し, 対象カテゴリーに応じて使用する識別器を選択している. また, [20] では, 検出にはタイプ 1 とタイプ 3 の手法を用いるが, アピランスの記述にはタイプ 1 の手法のみを用いる事で局所特徴が投影される特徴空間を同一のものにしている. [14] でも同様に, アピランスの記述にタイプ 1 の手法のみを用いている. しかし, 検出とアピランスの記述で用いる手法の局所特徴タイプが異なる場合, 検出された領域の画像特徴を十分に記述できるとはいえない. 検出とアピランスの記述に使用する手法の局所特徴タイプは一貫していることが重要である. [21] では, 検出とアピランスの記述が一貫したタイプ 1 とタイプ 3 の局所特徴を用いている. 各局所特徴タイプ別に識別器を構築しておき, あらかじめタイプ別に設定された重みを用いて各識別器の出力を統合している.

本研究では, 検出と記述の局所特徴タイプが一貫した 3 種類すべての局所特徴を用いる識別手法を提案する. 検出と記述の局所特徴タイプが一貫した局所特徴を組み合わせることで, より幅広い記述を可能にし, 識別性能の向上を図る. 識別器を各タイプごとに構築し識別結果を統合する.

本手法で扱う識別器は Hillel の手法 [11] をベースとする．各タイプごとに識別器を学習し，その識別器の出力を統合することで全局所特徴タイプを用いた識別を行う．本論文では，タイプ 1 の局所特徴には Hillel の手法で使われている KB detector と DCT を用いる．また，タイプ 2 とタイプ 3 には，3.2 と 3.3 で述べる手法を用いる．

本論文の構成は以下のとおりである．まず 2. では本論文で扱う識別器と，その統合方法について説明する．次に 3. では本論文で用いる各タイプの局所特徴について述べる．4. では実験について述べる．最後に 5. で本論文をまとめる．

2. 提案手法

提案手法で用いる識別器と，識別器の出力を統合する方法について述べる．なお提案手法は，画像中に識別対象の物体が存在するかないかを判別する 2 クラス識別問題を対象とする．

2.1 識別器

提案手法で用いる識別器は Hillel の手法 [11] をベースとする．この識別手法は生成的アプローチである星座モデル [12] に，識別性能向上のため識別的アプローチである AdaBoost を融合した手法である．星座モデルは，物体の部位（例：テールランプ，ナンバープレート等）のアピアランス，部位の位置，部位のスケールを，確率モデルとしてモデル化する．各部位はガウス分布として記述される．AdaBoost は弱識別器と呼ばれる簡単な識別器を複数用いて決定境界を構築する．Hillel の手法では弱識別器を星座モデルの部位に対応させることで生成モデルと判別モデルの融合を行っている．

本論文で用いる識別器の出力値は以下の式で計算される．これは，弱識別器を単一のガウス分布で表現した以外は，学習方法を含め完全に Hillel の手法と同一な手法である．なお弱識別器の変更は，実装の簡略化のために行った．

$$R(I) = \text{sign}(f(I)) \quad (1)$$

$$f(I) = \max_C \left\{ \sum_{k=1}^N \alpha^k h^k(I, C) - \nu \right\} \quad (2)$$

$$h^k(I, C) = \max_{x \in F(I, C)} G(x | \mu^k, \Sigma^k) \quad (3)$$

$R(I) = \{+1, -1\}$ であり，出力値が +1 の場合，入力画像 I 中に識別器が学習した物体が存在することを表

し，-1 の場合は画像中に存在しないことを表す． h^k は画像 I に対する k 番目の弱識別器の出力値である．出力値は実数値である． α^k が k 番目の弱識別器の重み， N が弱識別器の数である． ν がしきい値である． C は画像中の物体の中心位置とスケールの候補である．学習及び識別処理において，各画像ごとに適切な C が選択される．この処理により，画像中の物体の位置とスケールの変化に対処する． $F(I, C)$ は入力画像 I から得られた局所特徴の， C を考慮した特徴ベクトルの集合を， $G(\cdot | \mu, \Sigma)$ は平均ベクトル μ ，共分散行列 Σ のガウス分布を表す．弱識別器数 N はあらかじめ与え， α と ν ，ガウス分布のパラメータが学習で求まる．

識別器の学習手順を図 2 に示す．識別器を構成する弱識別器は逐次的に学習される．各弱識別器の学習はこう配降下型学習で行われる．また同時に C も最適化される．学習のループは大きく分けて 2 種類ある．一つは，弱識別器を逐次的に学習する大きなループ (Loop A)，一つは各弱識別器におけるこう配降下型学習のループ (Loop B) である．Loop A における p は，現在のループで逐次学習されている弱識別器のインデックスを表す．また，Loop B₂ 内で，弱識別器のパラメータ $\theta (= \{\mu, \Sigma\})$ の学習が，Loop B₃ 内で弱識別器の重み α が学習される．Loop B₁，B₂，B₃ の各ループ回数は，学習の制御パラメータとなる．詳しくは [11] を参照してほしい．ここで， i は学習画

Initialize $x^*, \theta, \alpha, C, w$.
Loop A $p = 1$ to N
 Loop B₁
 Loop B₂ (θ optimization)

$$\theta^p = \theta^p + \eta \sum_{i=1}^M w_i y_i \frac{dG(x_{i,p}^* | \theta^p)}{d\theta} \quad (4)$$

$$x_{i,p}^* = \operatorname{argmax}_{x \in F(I_i, C_i)} G(x | \theta^p) \quad (5)$$

 Loop B₃ (α optimization)

$$\alpha^p = \alpha^p + \eta \sum_{i=1}^M w_i y_i h^p(I_i, C_i) \quad (6)$$

$$[f_i^0, C_i] = \max_C \operatorname{argmax}_C \sum_{k=1}^p \alpha^k h^k(I_i, C_i) \quad (7)$$

$$\nu = \frac{1}{2} \log \left(\frac{\sum_{i: y_i = -1}^M \exp(f_i^0)}{\sum_{i: y_i = +1}^M \exp(-f_i^0)} \right) \quad (8)$$

$$w_i = \exp(-y_i \cdot (f_i^0 - \nu)) \quad (9)$$

図 2 識別器の学習手順

Fig. 2 Training algorithm for the classifier.

像のインデックスを、 M は学習画像枚数を表す。 y_i ($= \{+1, -1\}$) は各学習画像のラベル (物体が存在するかしないか) を、 w_i は学習処理内における各学習画像の重みを表す。また、 η はこう配降下型学習における更新パラメータである。

2.2 局所特徴の統合方法

各局所特徴タイプを統合する方法について述べる。各タイプ別に 2.1 で述べた識別器を学習し、各識別器の出力 $f(I)$ の和により、それぞれの識別結果を統合する。

統合された最終的な出力は、 l を局所特徴の各タイプのインデックスとすると、以下の式で算出される。

$$R_{comb}(I) = \text{sign} \left\{ \sum_{l=1}^3 f_l(I) \right\} \quad (10)$$

$f(I)$ は正負の実数値をとり、出力が信頼できるほど、その絶対値は大きくなる。したがって信頼できる局所特徴タイプほど、最終的な出力に対して支配的になる。識別結果のこの統合方法は、Boosting に代表されるモデル結合の変形と考えることができる。

3. 本論文で用いる各タイプの局所特徴

本章では、本論文で用いるタイプ 1 とタイプ 2、タイプ 3 の局所特徴の抽出法について説明する。

3.1 KB detector + DCT (タイプ 1)

タイプ 1 の局所特徴には、Hillel の論文中 [11] で使われている手法を用いる。検出に KB detector を、記述に DCT を用いる。まず KB detector により、タイプ 1 の領域を検出する。次に、得られた領域を入力画像から切り出し、それらすべてを同一サイズに正規化したあと、DCT で計算される直流を含まない最初の 20 個の係数を用いて記述する。局所特徴を記述する特徴量は、DCT 係数の 20 個の値、位置 (X, Y) を表す 2 個の値、領域のサイズを表す 1 個の値の 23 次元となる。なお X, Y 、領域のサイズは、物体の中心位置とスケールの候補 C を考慮した相対的な値となる。

3.2 一様領域局所特徴 (タイプ 2)

この局所特徴は、色が一様な楕円領域を表現する。色の局所的な一様性を表現するモデルのパラメータを画像上で推定することで抽出される。推定を開始する初期位置を、画像全体を走査するように少しずつずらして推定することで、画像全体に対して色が一様な領域を抽出する。

使用するモデルを図 3 に示す。 x 軸、 y 軸は画像の

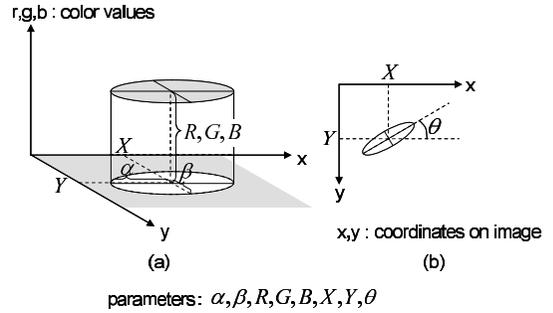


図 3 色が一様な領域を表現するモデル
Fig. 3 A model which describes a uniform color area.

横と縦の座標を表す。 r, g, b 軸は画像中のピクセルの色の R, G, B 値を表している。図の簡略化のため、 r, g, b 軸をまとめて一軸で表現している。図中の楕円は色が一様な領域を表現している。また、図 3 (b) は図 3 (a) を真上から見た図である。モデルのパラメータは 8 個あり、 α, β は楕円の長半径、短半径を、 R, G, B は楕円領域が表現する一様な色の値を、 X, Y は画像中の楕円領域の位置を、 θ は楕円の傾きを表す。

次に、モデルパラメータの推定法について説明する。モデルが楕円領域として表現する一様領域と、モデルが表す楕円領域中の画像領域との一致度が最大になるようにモデルの当てはめを行う。各パラメータに初期値を設定し、評価値が増加する方向にパラメータを更新し、評価値の増加が収束したらパラメータの推定を終了する。パラメータ R, G, B の初期値は推定を開始する初期位置のピクセルの R, G, B 値とする。 α, β の初期値は十分小さな値とする。 θ は 0 度とする。推定処理時の楕円領域サイズは、推定の初期では小さく、推定処理の進行とともに大きくなる。評価値の収束とともに楕円領域の拡大も収束する。

モデル M が楕円領域として表す一様領域と、モデル M の一様領域に相当する画像 I における画像領域との一致度を表す評価値について述べる。入力画像 I の位置 p のピクセルの R, G, B 値とモデル M のパラメータ R, G, B との距離 d は、それぞれのベクトル表現を $\text{RGB}_{I,p}, \text{RGB}_M$ とすると以下のように書ける。

$$d(I, M, p) = |\text{RGB}_{I,p} - \text{RGB}_M| \quad (11)$$

このとき、評価値の計算式 $f(I, M)$ を以下のように定義する。

$$f(I, M) = \sum_{p \in \text{ellipse}(M)} v(I, M, p) \quad (12)$$

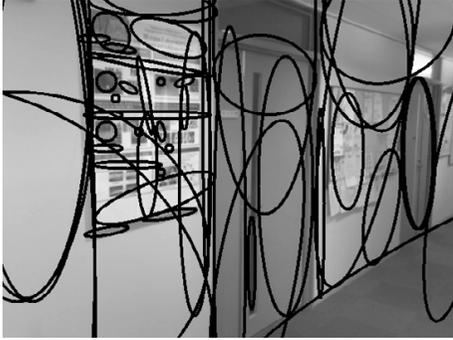


図 4 一様領域局所特徴の検出例
Fig. 4 Detection example of uniform color areas.

$$v(I, M, p) = \begin{cases} 1 & (d(I, M, p) < \epsilon) \\ -1 & (d(I, M, p) \geq \epsilon) \end{cases} \quad (13)$$

つまり評価値は、モデル M が表す楕円領域内に含まれるすべてのピクセルの位置 $ellipse(M)$ に対して、距離 d がしきい値 ϵ 未満のピクセルをカウントした値となる。ただし距離 d がしきい値 ϵ 以上のピクセルに対してはペナルティとして逆に評価値を減らす。なお、しきい値 ϵ は経験的に設定する。なお図 4 に一様領域局所特徴の検出例を示す。

得られた局所特徴は、モデルのパラメータを特徴量として表現する。ただし、 θ に関しては、 -90 度と $+90$ 度の間の数値の連続性を保持したまま、角度を特徴量として表現するため、以下で述べる変換を施した。

$$\begin{cases} a_1 = \cos(2\theta) \\ a_2 = \sin(2\theta) \end{cases} \quad (14)$$

以上を合わせて、特徴量は、 $\alpha, \beta, R, G, B, X, Y, a_1, a_2$ の 9 次元となる。なお α, β, X, Y は、物体の中心位置とスケールの候補 C を考慮した相対的な値となる。

3.3 部分エッジ局所特徴 (タイプ 3)

この局所特徴は、設定した長さの直線に近いエッジ線として画像中から検出され、検出されたエッジ線の角度と位置の値で記述される。

まず、入力画像からエッジ線を抽出する。本研究では文献 [22] にある手法を用いたが、Canny 法など他の手法を用いてもよい。次に、得られたエッジ線からランダムに点を選択し、その点の周りのエッジ線を局所特徴の候補とする。候補となったエッジ線の中には図 6 のような、角度による表現に適さない形状のも



図 5 直線に近いと判断されたエッジ線
Fig. 5 Edge lines judged to be nearly a straight line.



図 6 直線に近くないと判断されたエッジ線
Fig. 6 Edge lines judged to be not a straight line.



図 7 部分エッジ局所特徴の検出例
Fig. 7 Detection example of the partial edges.

のも含まれる。それらを排除するため、得られた局所特徴の候補の直線度合を判断する。

具体的には、選択した点周りの小領域 (例: $11 \text{ pixel} \times 11 \text{ pixel}$) を切り出し、選択した点が含まれるエッジ線のみを含む 2 値小画像をラベリング処理を用いて生成する (図 5, 図 6)。生成された 2 値小画像に対し主成分分析を行うことでエッジ線の直線度合を判断する。2 値小画像中のピクセルの座標 (X, Y) をパラメータ空間として、2 値小画像中の選択した点が含まれるエッジ線の各ピクセル (図 5, 図 6 中の白ピクセル) をサンプル点とし主成分分析を行う。主成分分析により得られる二つの固有値からエッジ線の直線度合を判断する。 $\frac{\text{第 1 固有値}}{\text{第 2 固有値}}$ (各固有値は各主成分方向の分散に対応) の値が十分大きなエッジ線を直線に近いエッジ線と判断し、最終的に得られた局所特徴とする。直線に近いと判断されたエッジ線の例を図 5 に、直線に近くないと判断されたエッジ線の例を図 6 に示す。主成分分析により得られた第 1 主成分の角度の値を、局所特徴を記述する角度の値とする。また、選択した点の位置を、局所特徴を記述する位置の値とする。以上の処理を任意の個数局所特徴が得られるまで繰り返す。なお局所特徴の候補の選択は、上述したようなランダムな選択ではなく、エッジ線上から一定間隔ごとで選択してもよい。図 7 に、部分エッジ局所特徴の

検出例を示す。

局所特徴を表現する特徴量は角度と位置である。角度は一樣領域局所特徴と同じように a_1, a_2 で表現する。したがって特徴量は a_1, a_2, X, Y の 4 次元となる。なお X, Y は、物体の中心位置とスケールの候補 C を考慮した相対的な値となる。

4. 実験

まず、3 種類の局所特徴の統合による識別性能の改善について検証する。また、既存手法との参考比較を行う。

4.1 実験条件

学習と検証用の画像データセットとして Caltech database [12] を、また、より難易度が高いとされる画像データセットとして Graz-02 [23] を用いる。

Caltech dataset は 4 カテゴリーの物体画像と 2 種類の背景画像で構成される。「物体画像」には単一カテゴリーの物体のみが含まれ、「背景画像」には一切の物体が含まれない。画像例を図 8 に示し、データの内訳を表 2 に示す。画像は大まかに向きと画像中の位置がそろえられているが、それぞれの画像で物体の見目は異なる。また Cars Rear と Motorbikes は画像中の物体の大きさについても異なる。

Graz-02 は 3 カテゴリーの物体画像と 1 種類の背景画像で構成される。図 9 に画像例を示す。また、データの内訳を表 3 に示す。すべての種類の物体において、それぞれの画像の物体の見目が大きく異なり、また物体の向きもそろえられていない。また、画像中の物体の位置や大きさのパリエーションも幅広い。したがって Caltech database よりも難易度の高いデータセットと考えられる。

本論文で対象としているタスクは、画像中に対象カテゴリーの物体が存在するかしないかを判別する 2 クラス識別問題である。識別処理は、カテゴリー別に行う。物体が存在する画像として対象カテゴリーの「物体画像」を、物体が存在しない画像として対象カテゴリーの「背景画像」を用いて、

識別率

$$= \frac{\text{「物体画像」が「背景画像」かを正しく識別できた画像枚数}}{\text{「物体画像」枚数} + \text{「背景画像」枚数}}$$

(15)

をカテゴリー別に計算する。その平均値で手法を評価する。カテゴリー別に、物体画像と背景画像の半分を

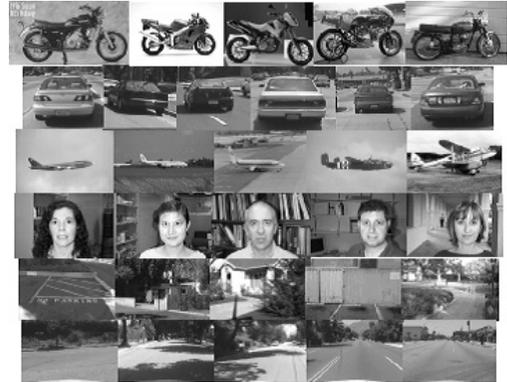


図 8 Caltech database の画像例 (上から Motorbikes, Car Rear, Airplanes, Faces, 背景画像 Airplanes・Motorbikes・Faces 用, 背景画像 Cars Rear 用)
Fig. 8 Example images from the Caltech database (From the top, Motorbikes, Car Rear, Airplanes, Faces, Background images for Airplanes, Motorbikes, and Faces, Background images for Car Rear).

表 2 Caltech database の内訳
Table 2 Details of the Caltech database.

Type of image		Num.
Object image	Airplanes	1074
	Cars Rear	1155
	Motorbikes	826
	Faces	450
Back ground image	For Airplanes・Motorbikes・Faces	900
	For Cars Rear	1370



図 9 Graz-02 の画像例 (上から Bikes, Persons, Cars, 背景画像 Bikes・Persons・Cars 用)
Fig. 9 Example images from the Graz-02 dataset (From the top, Bikes, Persons, Cars, and Background images for Bikes, Persons, Cars).

表 3 Graz-02 の内訳
Table 3 Details of the Graz-02 dataset.

Type of image		Num.
Object image	Bikes	365
	Persons	311
	Cars	420
Back ground image	For Bikes・Persons・Cars	380

表 4 Caltech database における用いる局所特徴タイプの違いによる識別率 (%) の変化 (1, 2, 3 はそれぞれ, タイプ 1, タイプ 2, タイプ 3 を表す)

Table 4 Recognition rates comparison on difference of used local feature types for Caltech database (%) (1, 2, 3 show type1, type2 and type3).

Cat. name	1 & 2 & 3	1 & 2	1 & 3	2 & 3	1	2	3
Motorbikes	99.42	98.72	98.61	99.30	94.20	98.95	97.91
Cars Rear	98.18	97.47	97.23	96.52	95.49	95.49	89.23
Airplanes	99.80	99.70	98.98	99.80	98.58	99.19	99.19
Faces	100.00	100.00	100.00	100.00	100.00	100.00	99.85
Average	99.35	98.97	98.70	98.90	97.07	98.41	96.54

表 5 Graz-02 における用いる局所特徴タイプの違いによる識別率 (%) の変化 (1, 2, 3 はそれぞれ, タイプ 1, タイプ 2, タイプ 3 を表す)

Table 5 Recognition rates comparison on difference of used local feature types for Graz-02 (%) (1, 2, 3 show type1, type2 and type3).

Cat. name	1 & 2 & 3	1 & 2	1 & 3	2 & 3	1	2	3
Bikes	78.55	77.75	78.02	75.07	77.75	61.66	75.34
Persons	79.19	77.46	77.46	76.88	76.30	73.99	77.75
Cars	78.00	68.25	70.00	70.25	68.75	68.25	64.00
Average	76.58	74.49	75.16	74.07	74.27	67.97	72.36

学習に使い, 残りを検証に用いた.

KB detector (タイプ 1) で検出した領域は $11 \text{ pixel} \times 11 \text{ pixel}$ に正規化した. 一様領域局所特徴 (タイプ 2) のしきい値 ϵ は 40 とした. また部分エッジ局所特徴 (タイプ 3) において, 切り出す小領域のサイズは $11 \text{ pixel} \times 11 \text{ pixel}$ とし, $\frac{\text{第 1 固有値}}{\text{第 2 固有値}}$ の値が 20 未満のエッジ線を直線に近くないと判断しランダムに 500 個の局所特徴を抽出した. 各タイプの識別器における弱識別器数 N はすべて, Hillel の論文 [11] と同様に 50 とした. また, 他の学習パラメータもすべて Hillel の論文 [11] と同様の値を用いた.

4.2 3 種類すべてのタイプを用いる有効性

3 種類のタイプを用いた場合と, 各タイプを単一で用いた場合及び各 2 種類を用いた場合とを比較し, 3 種類すべてを用いることの有効性を調査した. Caltech database と Graz-02 それぞれのデータセットにおいて, 3 種類すべて, 任意の 2 種類, 1 種類のみを用いた場合の識別率を表 4 と表 5 に示す.

結果は, 両方の画像データセットにおいて, 3 種類すべてを用いた場合が最も識別性能が高かった. この結果から, 3 種類すべての局所特徴を使用することの有効性が示された.

また, 表 4 と表 5 からカテゴリごとに識別性能が最も高い局所特徴タイプが異なっていることが分かる. 例えば表 4 の Caltech database の場合, Motorbikes ではタイプ 2 が, Cars Rears ではタイプ 1 と 2 が, Airplanes ではタイプ 2 と 3 が, Faces ではタイプ 1 と 2 が識別性能の最も高い局所特徴タイプである. こ

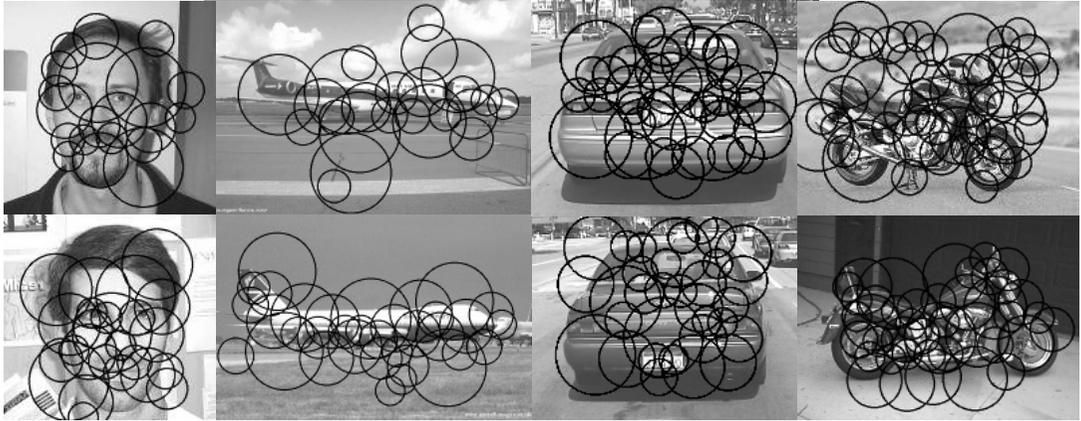
のことからも, カテゴリごとに有効な局所特徴タイプが異なり, 三つすべての局所特徴タイプを用いることの有効性がうかがえる.

また, 各タイプが記述する領域の違いを具体的に示すため, 実験で用いた各タイプの局所特徴の検出例を図 10 に示す. タイプ 1 である KB detector は, 例えば Car Rears のナンバープレートやテールランプ付近など特徴的なテクスチャを含む領域を, タイプ 2 である一様領域局所特徴は, 例えば Faces のおでこや頬など画像の性質が一般的な画像領域を, タイプ 3 である部分エッジ局所特徴は, 例えば Airplanes の輪郭など, 主に物体の形状を記述していることがこれら検出例から分かる.

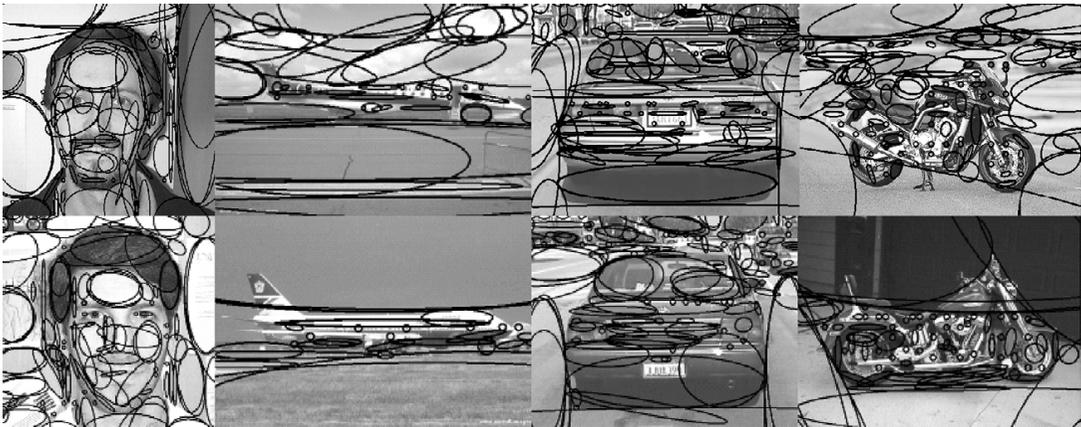
4.3 既存手法との参考比較

本論文で用いたデータセットと同じデータセットを使用した研究が存在するため, 提案する 3 種類の局所特徴タイプを用いた識別手法と既存手法の識別率を参考として比較する. 各手法で用いられている局所特徴のタイプと識別率を, 表 6, 表 7 に示す^(注1). ただしそれぞれの手法で実験条件が異なるため, この比較は 3 種類すべてのタイプを利用することの優位性を, 既

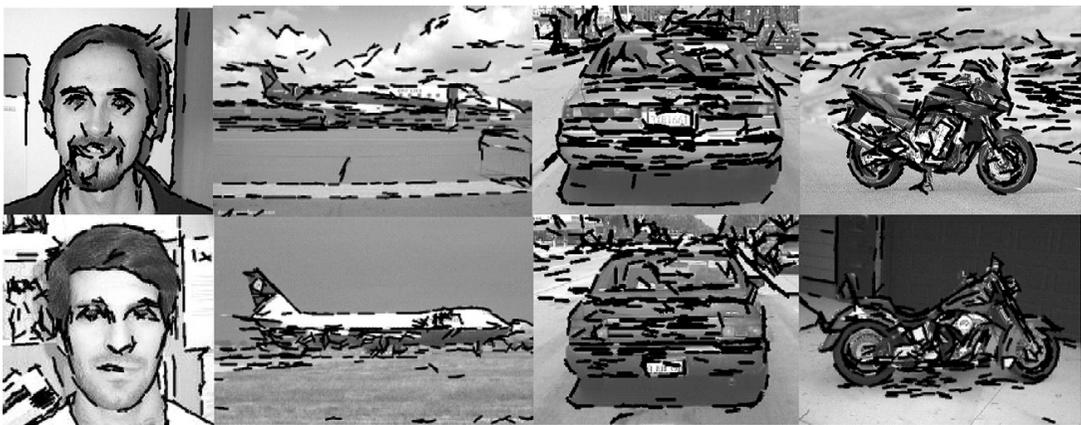
(注1): 各論文は ROC equal error rate で評価している. この計算手順は, まずしきい値等パラメータを変化させ複数の識別率を求め, 最後に FPR (false positive rate) と FNR (false negative rate) が等しい識別率を選択する. Hillel の論文 [11] では, この作業を, 評価時にしきい値を決定する方法ととらえており, 一方 Hillel の方法では, しきい値 ν は最適値が自動的に学習されるので, この作業でしきい値を決定する必要はなく, 学習で得られた ν で計算した識別率で評価を行っている. 本論文でも, それに従い, 学習で得られた ν による識別率で評価を行う.



(a) KB detector(タイプ1) による Caltech database における検出例
(a) Detection examples of the local feature by KB detector (type 1) on Caltech database.



(b) 一様領域局所特徴 (タイプ2) の Caltech database における検出例
(b) Detection examples of the uniform area local feature (type 2) on Caltech database.



(c) 部分エッジ局所特徴 (タイプ3) の Caltech database における検出例
(c) Detection examples of the partial edge local feature (type 3) on Caltech database.

図 10 各局所特徴タイプが記述する領域の比較

Fig. 10 Comparison of described areas by each local feature type.

表 6 Caltech database における既存研究との識別率の参考比較 (%)
Table 6 Comparison of recognition rate with previous works (%).

	Our method	Hillel [11]	Fergus [12]	Opelt [18]	Opelt [21]
Type of local feature	1&2&3	1	1	3	1&3
Motorbikes	99.42	95.1	93.3	96.8	100.0
Car Rearrs	98.18	99.4	90.3	99.5	99.5
Airplanes	99.80	93.3	93.0	97.4	97.1
Faces	100.00	93.7	96.3	98.1	99.7
Average	99.35	95.38	93.25	97.95	99.07

表 7 Graz-02 における既存研究との識別率の参考比較 (%)
Table 7 Comparison of recognition rate with previous works (%).

	Our method	Opelt [23]
Type of local feature	1&2&3	1&2
Bikes	78.55	77.8
Persons	79.19	81.2
Cars	72.00	70.5
Average	76.58	76.5

存手法に対して示しているものではなく、あくまで参考比較であることに注意してほしい。

5. む す び

本研究では、これまで提案されてきた局所特徴を 3 種類に分類し、各タイプの局所特徴（特徴的なテクスチャに着目する局所特徴、画像領域の一様性に着目する局所特徴、エッジ情報に着目する局所特徴）をすべて組み合わせた識別手法を提案した。実験ではすべてのタイプの局所特徴を用いることが識別性能の改善につながることを示した。

本論文では、タイプ 2 とタイプ 3 として本論文で述べた手法を使ったが、これらに対する手法は他の既存手法も用いることができる。今後の課題として、各タイプで用いる手法の組合せの最適化がある。同一タイプに分類される手法でも個々の手法の特性は異なり、組合せに応じて識別性能の変動が生じる。各タイプで用いる手法の組合せを最適化することで識別性能の更なる改善を図る。

文 献

- [1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol.60, no.2, pp.91-110, Jan. 2004.
- [2] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," *Proc. European Conf. on Computer Vision*, pp.128-142, 2002.
- [3] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol.45, no.2, pp.83-105, Nov. 2001.
- [4] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," *Proc. Third Int. Conf. on Visual Information Systems*, 1999.
- [5] Y. Deng and B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.23, no.8, pp.800-810, 2001.
- [6] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *Proc. British Machine Vision Conf.*, pp.384-396, 2002.
- [7] J. Garding and T. Lindeberg, "Direct computation of shape cues using scale-adapted spatial derivative operators," *Int. J. Comput. Vis.*, vol.17, no.2, pp.163-191, Feb. 1996.
- [8] J. Malik, S. Belongie, J. Shi, and T. Leung, "Textons, contours and regions: Cue integration in image segmentation," *Proc. IEEE Int. Conf. on Computer Vision*, vol.2, pp.918-925, 1999.
- [9] 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.
- [10] 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.
- [11] A.B. Hillel, T. Hertz, and D. Weinshall, "Efficient learning of relational object class models," *Proc. IEEE Int. Conf. on Computer Vision*, vol.2, pp.1762-1769, 2005.
- [12] 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.
- [13] C. Pantofaru, G. Dorko, C. Schmid, and M. Hebert, "Combining regions and patches for object class localization," *Proc. Beyond Patches Workshop in Con-*

- junction with the IEEE Conf. on Computer Vision and Pattern Recognition, pp.23–30, 2006.
- [14] J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” Proc. IEEE Int. Conf. on Computer Vision, pp.1470–1477, 2003.
- [15] S. Lazebnik, C. Schmid, and J. Ponce, “Semi-local affine parts for object recognition,” Proc. British Machine Vision Conf., vol.2, pp.779–788, 2004.
- [16] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, “Groups of adjacent contour segments for object detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.1, pp.36–51, 2008.
- [17] M. Leordeanu, M. Hebert, and R. Sukthankar, “Beyond local appearance: Category recognition from pairwise interactions of simple features,” Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.1–8, 2007.
- [18] A. Opelt, A. Pinz, and A. Zisserman, “A boundary-fragment-model for object detection,” Proc. European Conf. on Computer Vision, vol.2, pp.575–588, 2006.
- [19] R. Fergus, P. Perona, and A. Zisserman, “A visual category filter for google images,” Proc. European Conf. on Computer Vision, vol.1, pp.242–255, 2004.
- [20] 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.
- [21] A. Opelt, A. Pinz, and A. Zisserman, “Fusing shape and appearance information for object category detection,” Proc. British Machine Vision Conf., 2006.
- [22] P. Meer and B. Georgescu, “Edge detection with embedded confidence,” IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.12, pp.1351–1365, Dec. 2001.
- [23] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, “Generic object recognition with boosting,” IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.3, pp.416–431, March 2006.

(平成 20 年 8 月 13 日受付, 12 月 1 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課程前期課程了。現在, 同大学院情報科学研究科博士課程後期課程在学中。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。



矢野 良和

2003 名古屋大学大学院工学研究科博士課程後期課程了。2003 同大学助手。2006 愛知工業大学講師, 現在に至る。主として画像処理, 感情情報処理, ソフトコンピューティングに関する研究に従事。電気学会, 情報処理学会, 日本知能情報フジィ学会, 日本音響学会各会員。



大熊 繁

1972 名古屋大学大学院工学研究科修士課程電気・電子工学専攻了。1974 年 6 月 ケース・ウェスタン・リザーブ大学大学院修士課程システム工学専攻了。1977 年 3 月 名古屋大学大学院工学研究科博士課程電気・電子工学専攻単位取得退学。同年 4 月同大学助手, 1979 年 4 月同講師, 1984 年 11 月同助教授, 1990 年 12 月同教授, 現在に至る。工博。主として, ロボティクス, パワーエレクトロニクス, ソフトコンピュータに関する研究に従事。電気学会, 計測自動制御学会, ロボット学会, IEEE, パワーエレクトロニクス研究会会員。IEEE IECON'92 Best paper award, 日本精密工学会論文賞, 電気学会論文賞 2 回受賞。



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, 18, 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．