

ロボットビジョンのための教師なし学習による 特徴点選択とカテゴリ分類

塚田 真敬[†] 内海 祐哉[†] 間所 洋和[†] 佐藤 和人[†]

[†] 秋田県立大学大学院システム科学技術研究科 〒015-0055 秋田県由利本荘市土谷字海老ノ口 84-4
E-mail: †{m11a013,m12a005,madokoro,ksato}@akita-pu.ac.jp

あらまし 本論文では、教師なし学習に基づく特徴点の選択法と、事前にカテゴリ数の設定を必要としない教師なしカテゴリ分類法を提案する。本手法は (1) Scale-Invariant Feature Transform(SIFT) を用いた特徴点の抽出および特徴量の記述 (2) One Class-Support Vector Machines(OC-SVMs) を用いた SIFT 特徴点の選択 (3) Self-Organizing Maps(SOMs) を用いた全 SIFT 記述子ベクトルによる visual words の作成および選択した特徴量による各画像のヒストグラムの作成 (4) Adaptive Resonance Theory-2(ART-2) を用いたラベルの生成 (5) ART-2 のラベルを用いた Counter Propagation Networks(CPNs) のラベリングによるカテゴリマップの作成の手順で構成する。Caltech-256 を用いた画像分類実験と、ロボットの移動とともに得られる視野画像列を用いたカテゴリ分類実験から、入力画像の系列性を保持しつつカテゴリの空間関係が可視化できる本手法の特性と、オブジェクトの見え方の変化に対するカテゴリ分類における本手法の有効性を示す。

キーワード ART-2, CPNs, 自己組織化マップ, SIFT, OC-SVMs, 教師なしカテゴリ分類, ロボットビジョン

Unsupervised Learning for Selection of Feature Points and Category Classification Using Robot Vision

Masahiro TSUKADA[†], Yuya UTSUMI[†], Hirokazu MADOKORO[†], and Kazuhito SATO[†]

[†] Graduate School of Systems Science and Technology, Akita Prefectural University 84-4 Aza Ebinokuchi
Tsuchiya, Yurihonjo City, 015-0051 Japan

E-mail: †{m11a013,m12a005,madokoro,ksato}@akita-pu.ac.jp

Abstract This paper presents an unsupervised learning-based method for selection of feature points and object category classification without previous setting of the number of categories. Our method consists of the following procedures: 1) detecting feature points and description of features using a Scale-Invariant Feature Transform (SIFT), 2) selecting target feature points using One Class-Support Vector Machines (OC-SVMs), 3) generating visual words of all SIFT descriptors and histograms in each image of selected feature points using Self-Organizing Maps (SOMs), 4) forming labels using Adaptive Resonance Theory-2 (ART-2), and 5) creating and classifying categories on a category map of Counter Propagation Networks (CPNs) for visualizing spatial relations between categories. Classification results of static images using Caltech-256 object category dataset and dynamic images using time-series images obtained by a robot according to movements respectively portray that our method can visualize spatial relations of categories while maintaining time-series characteristics. Moreover, we show the effectiveness of our method for category classification of appearance changes of objects.

Key words ART-2, CPNs, SOMs, SIFT, OC-SVMs, Unsupervised Category Classification, Robot Vision

1. はじめに

近年、コンピュータビジョンの分野において一般物体認識が盛んに研究されている [1]。一般物体認識とは、制約のない実画像に対して計算機がその中に含まれる物体もしくはシーンを一般的な名称で認識することと定義される。つまり、画像中の物体もしくはシーンと意味カテ

ゴリ認識となる。また、ロボティクスの研究において、様々な環境に柔軟に適應する学習機能を有するロボットの實現には、世界像の獲得が求められている [2]。世界像とは、世界を折りたたんで脳の中に入れたようなものと定義される。つまり、世界像を生成するために、ロボットは視野画像列に含まれる物体やシーンをカテゴリに分類し、長期記憶として蓄積する必要がある。また、口

ロボットが移動する実環境は、カテゴリ数が一意に決められないため、カテゴリ数が未知な分類問題となる。そのため、追加的にカテゴリを生成しながら分類する必要がある。

本論文では移動ロボットへの応用を目的とした教師なしカテゴリ分類法を提案する。本手法の特徴は以下の4つとなる。1つ目は、One Class-Support Vector Machines (OC-SVMs) を用いることによって事前に境界情報を必要としない注目領域の絞り込みができる。2つ目は、時系列データに対して安定性と可塑性を保ちながらカテゴリの候補となるラベルを生成することができる。3つ目は、Adaptive Resonance Theory-2 (ART-2) により生成されるラベルを Counter Propagation Networks (CPNs) の教師信号として用いることにより、カテゴリマップの自動ラベリングが実現できる。4つ目は、CPNs の2次元マップ上に各カテゴリの空間的な関係を可視化し、カテゴリ内に含まれる多様性が表現できる点である。一般物体認識のアルゴリズムの性能比較のための標準的なベンチマークデータセットである Caltech-256 を用いた画像分類実験と、移動ロボットのカメラから得られる視野画像列を用いたカテゴリ分類実験によって本手法を評価する。

2. 関連研究

学習に基づくカテゴリ分類法は、教師ありカテゴリ分類法と教師なしカテゴリ分類法に大別される。教師ありカテゴリ分類法は、Ground-truth から抽出した教師信号を含む学習用データが必要となる。一方、教師なしカテゴリ分類法は、明示的な教師信号は必要とせず、分類カテゴリが未知な問題に対して自動的にカテゴリを探し出し、画像をカテゴリに分類する。近年では、様々な環境において柔軟に視覚情報を分類する技術として、教師なしカテゴリ分類法が注目され、活発に研究されている。

カテゴリ分類では、画像全体を対象とするのではなく、分類および認識対象となるオブジェクトが含まれる領域又は位置を検出する処理と組み合わせた様々な手法が提案されている。Barnard らは、領域に基づく手法として word-image-translation model を提案している [3]。彼らは、予め複数個のキーワードが付けられている画像を用いて、領域分割された画像への自動アノテーションを行った。鈴木らは、SVMs を用いた Bag-of-Features (BoF) における局所特徴の特徴選択法を提案している [4]。これは、SIFT 特徴量を SVMs により背景領域と対象領域の特徴に分類し BoF に利用する手法である。しかしながらこれらの手法は、事前に正解データを必要とするため、対象領域と背景領域を一意に決めることのできない実環境には適用できない。

教師なしカテゴリ分類法としては、Sivic らは BoF に the probabilistic Latent Semantic Analysis (pLSAs) と the Latent Dirichlet Allocation (LDA) の生成モデルを適

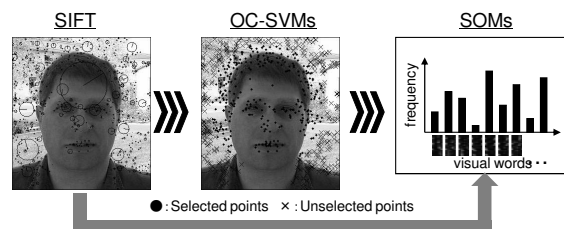


図1 本手法の Bag-of-Features

Fig.1 Procedures of our image representation method based on BoF.

用した教師なしカテゴリ分類法を提案している [5]。彼らは、各オブジェクトを複数のトピックの組み合わせとしてモデル化し、大量の画像からオブジェクトカテゴリとしてトピックを発見することを試みた。Zhu らは、Probabilistic Context-Free Grammars (PCFGs) と Markov Random Fields (MRFs) を組み合わせた生成モデルの Probabilistic Grammar-Markov Models (PGMMs) を取り入れた [6]。彼らは、この手法を用いて物体検出と教師なしカテゴリ分類のための物体カテゴリのモデルを生成した。さらに、この手法を発展させた Probabilistic Object Models (POMs) を提案し、オブジェクトの分類、セグメンテーション、認識を可能とした [7]。また、中村らは、視覚・聴覚・触覚のマルチモーダル情報を用いた教師なしカテゴリ分類法を提案している [8]。彼らは、ロボットの身体性を利用することで、人間の感覚に近い物体のカテゴリ分類を実現した。しかしながら、これらの手法は、分類カテゴリ数が既知でなければならないという制約がある。そのため、カテゴリ数が未知な実環境でのカテゴリ分類には適用できない。

3. 画像表現法

大量の画像から抽出した局所特徴量から代表的なパターンを visual words として作成し、それに関するヒストグラムにより特徴を表現する BoF は、一般物体認識の画像表現法において有効性が示されており、広く用いられている。本手法における BoF は、図1に示すように、画像全体の注目領域における SIFT 特徴量の選択に OC-SVMs を適用し、visual words の作成および選択した特徴量による各画像のヒストグラムの作成に Self-Organizing Maps (SOMs) を適用した。一般的に、SIFT 特徴点の密度は、前景領域が高く、背景領域が低いと考えられる。したがって本手法では、OC-SVMs を用いて前景領域となる注目領域の SIFT 特徴量を絞り込む。OC-SVMs は、密度関数の推定を行わずに、密度推定ができる教師なし学習の2値分類器であるため、境界を一意に決めることができない実環境にも適用可能である。以下に OC-SVMs, SOMs の詳細について記す。

3.1 OC-SVMs による特徴点の選択

OC-SVMs は、カーネルを用いた特徴空間への写像に

より、入力データの外れ値が特徴空間の原点近くに写像される性質を利用し、原点近くと他のデータを識別する超平面を設定する。本手法における OC-SVMs への入力データは、SIFT 特徴点の X, Y 座標である。超平面の位置は、パラメータ ν によって変わり、 ν は $0 < \nu \leq 1$ の範囲で、入力データの外れ値になるデータを制御する。入力特徴空間を 2 つに分けるために、識別関数により 2 値を計算する。

$$f(x) = \text{sign}(\omega^\top \Phi(x) - \rho) \quad (1)$$

ここで、 ω は線形識別器の重みベクトル、 Φ は高次元特徴空間上の点、 \top は内積、 ρ ($\rho \in R$) はマージンを示す。入力ベクトル x_i を高次元特徴空間に写像したときの結果を z_i とする。

$$\Phi : x_i \mapsto z_i \quad (2)$$

また、制約条件を以下のように設定する。

$$\begin{aligned} \omega^\top z_i &\geq \rho - \zeta_i, (i = 1, \dots, l) \\ \zeta_i &\geq 0, (i = 1, \dots, l), 0 < \nu \leq 1 \end{aligned} \quad (3)$$

ここで、 ζ は緩和変数ベクトルとする。制約条件のもとで、次の最適化問題を解く。

$$\begin{aligned} \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \zeta_i - \rho \\ \rightarrow \min \omega, \zeta, \text{ and } \rho \end{aligned} \quad (4)$$

式 (4) の最適化問題の解が、 $\rho \neq 0$ を満たす場合、OC-SVMs のパラメータ ν は、外れデータ数の割合の上限であり、サポートベクタ数の割合の下限である。

3.2 SOMs を用いた visual words の作成

本手法では、BoF における visual words の作成に一般的に用いられる k-means ではなく、SOMs を適用する。SOMs の学習フェーズは、入力データの位相構造を保ちながら、ネットワークの結合荷重を変更する。また、入力データに最も類似するユニットが発火し、その周辺のユニットが近傍領域を形成する。よって SOMs は、学習データの分布と類似する様々なデータを分類することができる。寺島らは、k-means より SOMs の方が誤認識率を最小とする教師なしクラス分類法であることを示している [9]。SOMs の学習アルゴリズムは、CPNs の入力層と Kohonen 層との間のアルゴリズムと同じである [10]。本手法では、SOMs の学習フェーズとなる visual words の作成には全 SIFT 特徴量を入力し、visual words に基づいたヒストグラムの作成には OC-SVMs で選択された SIFT 特徴量を入力する。

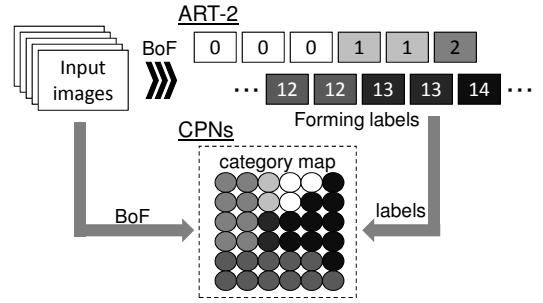


図 2 教師なしカテゴリ分類法

Fig. 2 Unsupervised category classification method.

4. 教師なしカテゴリ分類法

本手法におけるカテゴリ分類法は、図 2 に示すように、ART-2 の追加学習機能と、CPNs の自己写像特性を組み合わせ、教師なしカテゴリ分類法である。2. で述べたように、教師なしカテゴリ分類の主流を成す確率グラフモデルを用いた手法は、高い性能を有するものの、識別結果はグラフ構造に依存する。また、現状は実験者や開発者が経験に基づいて手動でグラフ構造を設計しており、自動でグラフを生成するツールも存在しているが高い性能は示されていない [11]。一方我々の提案する手法は、ART-2 と CPNs を組み合わせることにより、事前にカテゴリ数の設定を必要とせず、追加的にカテゴリの候補となるラベルが生成でき、それらの空間関係を可視化することが可能となる。移動ロボットの視野画像列の特徴は、時間の経過と共に変化するため、時系列に沿った学習ができる ART-2 の適用は、対象データが時系列変化である上でも有用と考えられる。カテゴリ分類結果は、CPNs の Kohonen 層ユニットにカテゴリマップとして表現される。以下に ART-2 と CPNs のアルゴリズムを記す。

4.1 ART-2 を用いたラベルの形成

ART には様々なタイプが存在し、本手法では、連続値入力が可能な ART-2 を用いる [12]。

ART2 のネットワークは、特徴表現の Field 1 (F1) と、カテゴリ表現の Field 2 (F2) から構成されている。F1 は、複数のサブレイヤ ($p_i, q_i, u_i, v_i, w_i, x_i$) から成り、入力データが各サブレイヤを遷移することによって短期記憶を実現する。また、短期記憶の効果によって、入力データ内のノイズが除去され、特徴が強調される。F2 には、位相の強弱により長期記憶としてカテゴリが形成される。本研究では、このカテゴリをラベルとして用いる。

F1 と F2 は p_i を介して接続されている。入力 $I_i(t)$ に対して F1 伝搬後の F2 の最大活性化ユニット T_j は、

$$T_j(t) = \max_j \left(\sum_i p_i(t) Z_{ij}(t) \right), \quad (5)$$

となる． T_j を基準として，トップダウン結合荷重 Z_{ji} とボトムアップ結合荷重 Z_{ij} は次式で更新される．

$$\frac{d}{dt} Z_{ji}(t) = d[p_i(t) - Z_{ji}(t)], \quad (6)$$

$$\frac{d}{dt} Z_{ij}(t) = d[p_i(t) - Z_{ij}(t)], \quad (7)$$

次に，ビジランス閾値 ρ を用いて，カテゴリに属するかを判定する．

$$r_i(t) = \frac{u_i(t) + cp_i(t)}{e + \|u\| + \|cp\|}, \quad \frac{\rho}{e + \|r\|} > 1. \quad (8)$$

判定が成立する場合は，選択されたユニットをリセットして再探索する．不成立の場合は，F1 層内の変化率が小さくなるまで伝搬と結合荷重の更新を繰り返す．

4.2 CPNs を用いたカテゴリマップの生成

CPNs は，競合学習と近傍学習に基づきパターンを特定のカテゴリに分類する [13]．本手法は，Grossberg 層のユニットに与える教師データを ART-2 のラベルとすることで，ラベリング処理を自動化している．

CPNs の学習アルゴリズムを以下に記す．

(1) $u_{n,m}^i(t)$ は，時刻 t における，入力層ユニット i ($i = 1, \dots, I$) から，Kohonen 層ユニット (n, m) ($n = 1, \dots, N, m = 1, \dots, M$) への結合荷重とする． $v_{n,m}^j(t)$ は，時刻 t における，Grossberg 層ユニット j から，Kohonen 層ユニット (n, m) への結合荷重とする．これらの結合荷重は，ランダムに初期化される． $x_i(t)$ は，時刻 t における入力層ユニット i に提示される学習データである． $x_i(t)$ と $u_{n,m}^i(t)$ の間のユークリッド距離 $d_{n,m}$ は次式で計算される．

$$d_{n,m} = \sqrt{\sum_{i=1}^I (x_i(t) - u_{n,m}^i(t))^2}. \quad (9)$$

(2) $d_{n,m}$ が最小となるユニットが，勝者ユニット c として定義される．

$$c = \operatorname{argmin}(d_{n,m}). \quad (10)$$

(3) $N_c(t)$ は，勝者ユニット c の近傍領域である． $N_c(t)$ の内部の結合荷重 $u_{n,m}^i(t)$ は，Kohonen の学習アルゴリズムを用いて更新される．

$$u_{n,m}^i(t+1) = u_{n,m}^i(t) + \alpha(t)(x_i(t) - u_{n,m}^i(t)). \quad (11)$$

(4) $N_c(t)$ の内部の結合荷重 $v_{n,m}^j(t)$ は，Grossberg のアウトスター学習アルゴリズムで更新される．

$$v_{n,m}^j(t+1) = v_{n,m}^j(t) + \beta(t)(t_j(t) - v_{n,m}^j(t)). \quad (12)$$

ここで， $t_j(t)$ は，Grossberg 層に提示される教師信号である． $\alpha(t)$ と $\beta(t)$ は，学習率係数であり，学習の進行とともに減少する．CPNs の学習は，以上のステップを事前に設定した学習回数だけ繰り返す．

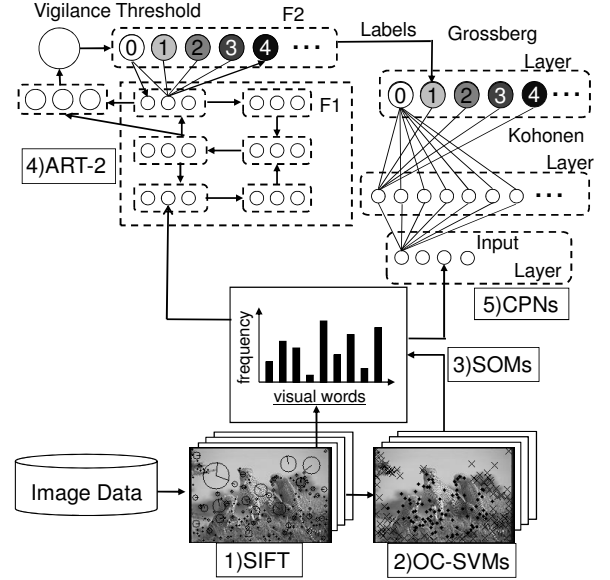


図3 提案手法のネットワーク構成

Fig. 3 The whole architecture of our method.

5. 提案手法の全体構成

本研究で提案するカテゴリ分類法のネットワーク構成を図3に示す．処理手順は，

- (1) SIFT を用いた特徴点の抽出と特徴量の記述，
- (2) OC-SVMs を用いた SIFT 特徴量の選択，
- (3) SOMs を用いた全 SIFT 記述子ベクトルによる visual words の作成，および選択した特徴量による各画像のヒストグラムの作成，
- (4) ART-2 を用いたラベルの生成，
- (5) ART-2 のラベルを用いた CPNs のラベリングによるカテゴリマップの作成となる．

画像表現法に当たる (1)~(3) は，BoF 表現をベースとしている．SIFT 特徴量の選択として OC-SVMs を適用し，画像中のオブジェクトなどの注目領域を絞り込む．visual words の作成には SOMs を用いる．さらに，ART-2 と CPNs の組み合わせが，大量の画像に自動でラベルを与えて各カテゴリに分類し，教師なしカテゴリ分類を可能とする．各実験で用いたパラメータは，OC-SVMs の ν が 0.5，ART-2 の θ が 0.1， ρ が 0.920，CPNs の $\alpha(t)$ および $\beta(t)$ が 0.5，学習回数が 10,000 回とした．

6. Caltech-256 を用いた実験

本節では，Caltech-256 を用いた画像分類実験を行う．Caltech-256 に含まれる各カテゴリには，時間的な関係性が存在しないため，本実験は静的画像のカテゴリ分類と位置付ける．本実験では，256 種類のカテゴリの中から，データ数の多い上位 20 カテゴリを用いた．以下に特徴点の選択結果および 5，20 カテゴリ分類の認識率を示す．

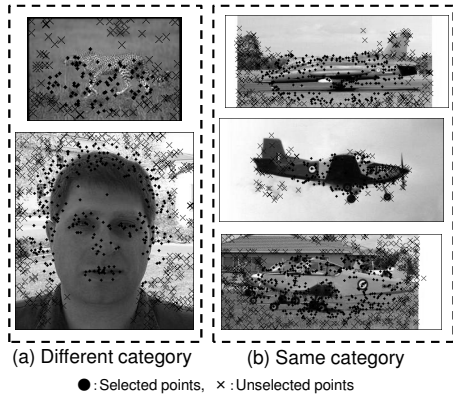


図 4 caltech-256 における SIFT 特徴点の選択結果

Fig. 4 Results of selected SIFT feature on two sample images in different category and three sample images in same category of Caltech-256.

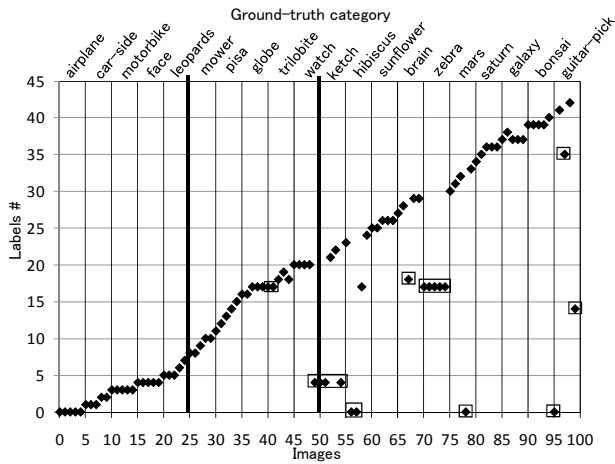


図 5 20 カテゴリ分類の ART-2 によるラベル形成の結果

Fig. 5 Results of formed labels using ART-2 at 10 and 20 categories.

6.1 特徴点選択及びラベル生成結果

OC-SVMs による SIFT 特徴点の選択結果を図 4 に示す．図 4(a) より，leopards と face カテゴリの画像に対して，認識対象オブジェクトの SIFT 特徴点を選択されている．また，図 4(b) に示す同一カテゴリの異なる画像に対しても airplane を特徴付ける翼付近の SIFT 特徴点を選択されている．ART-2 によるラベル生成結果を図 5 に示す．縦軸には形成したラベルを，横軸には画像を示す．図中の太線は 5 および 10 カテゴリまでの画像数を示す．8 カテゴリ目までは独立したラベルを形成できた．また，airplane，motorbike，face カテゴリは，1 種類のラベルが形成され，car-side と leopards カテゴリは複数のラベルが形成された．これらの結果から，各カテゴリ内でも背景やオブジェクトの形状および見え方が異なる画像に対して，OC-SVMs によって認識対象であるオブジェクトの SIFT 特徴点を選択し，ART-2 によって独立したラベルを形成できた．

また，Caltech-256 に含まれる 20 カテゴリ分類を用い

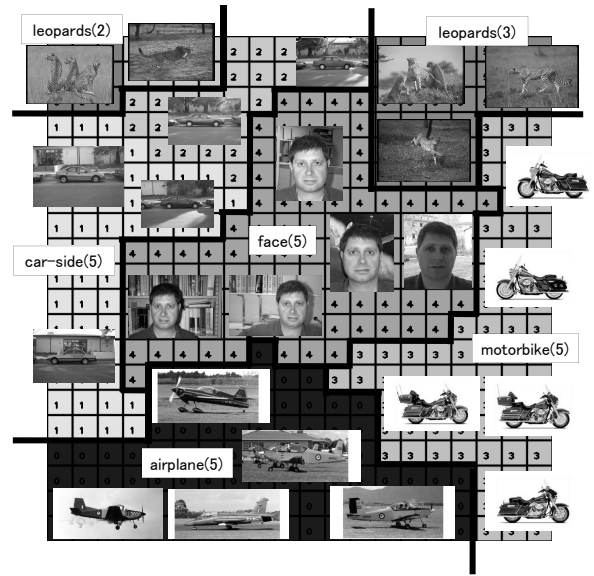


図 6 5 カテゴリ分類の CPNs によるカテゴリマップの結果
Fig. 6 Result of category map using CPNs of five categories.

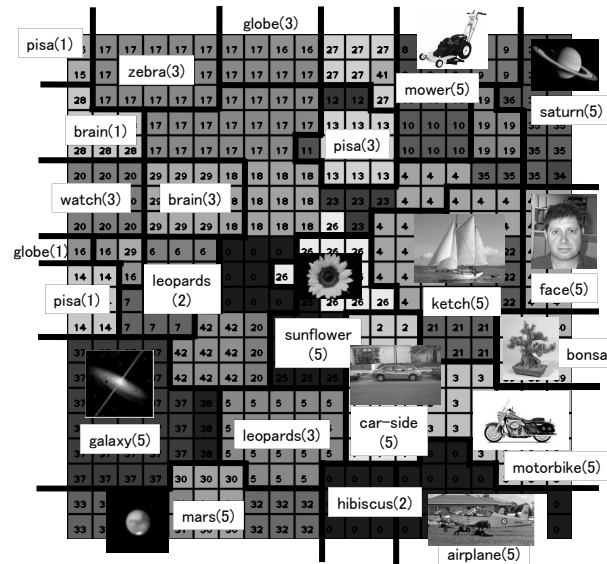


図 7 20 カテゴリ分類の CPNs によるカテゴリマップの結果
Fig. 7 Result of a category map using CPNs of 20 categories.

た実験における ART-2 によるラベル形成の結果を図 5 に示す．図中の四角で囲むカテゴリは，ラベルが混同した画像である．10 カテゴリでは 3 画像でラベルが混同したが，全てのカテゴリに独立したラベルを形成できた．また，20 カテゴリでは 16 画像でラベルが混同し，全ての画像でラベルが混同した zebra カテゴリを除いた 19 カテゴリに独立したラベルを形成できた．ketch，hibiscus，guitar-pick カテゴリでは，多くの画像でラベルが混同した．このように ART-2 は，10 カテゴリ分類まではラベルの混同を抑制できたが，それ以上ではラベルの混同が多く現れた．

表 1 Caltech-256 における学習とテストの認識率

Table 1 Recognition rates of learning and testing datasets used in Caltech-256.

	Learning	Testing
5 category	96%	76%
10 category	94%	44%
20 category	81%	50%

6.2 カテゴリ分類結果

Caltech-256 に含まれる airplane, car-side, motor-bike, face, leopards カテゴリのデータセットを用いた 5 カテゴリ分類実験におけるカテゴリマップの生成結果を図 6 に示す。カテゴリマップ上には、各ユニットに写像された画像と各カテゴリの写像領域を示す。図より、ART-2 で形成したラベルをもとに、CPNs が各カテゴリの画像を隣接したユニットに写像することで各カテゴリを形成できた。ART-2 で複数のラベルが形成された car-side と leopards カテゴリに着目すると、car-side カテゴリは隣接したユニットに写像されたが、leopards カテゴリは 2 つに分かれた。これは同一カテゴリにおいても、位相空間上では特徴が大きく異なるためと考えられる。

20 カテゴリ分類におけるカテゴリマップの生成結果を図 7 に示す。カテゴリマップ上に示すラベルは、Ground Truth に基づいて割り当てたカテゴリと、各カテゴリに対応付く画像枚数を示す。20 カテゴリ中 11 カテゴリは、全ての画像を隣接したユニットに写像できている。これらの画像は、ART-2 では複数の異なるラベルに分かれていたが、CPNs ではカテゴリ毎に独立した領域を形成できている。また、ART-2 でラベルが混同した globe, trilobite, zebra カテゴリに着目すると、CPNs では各カテゴリで隣接したユニットに写像できた。なお、カテゴリマップ上にカテゴリ名を示していない箇所には、複数のカテゴリの画像が写像されている。ここで、本手法の分類性能を定量的に評価するために、以下に示す認識率を求めた。

$$(\text{認識率}) = \frac{(\text{正解データ})}{(\text{全データ})} \times 100. \quad (13)$$

5, 10, 20 カテゴリにおける学習とテストの認識率を表 1 に示す。学習データに対する認識率は、10 カテゴリまでは 90%以上、20 カテゴリで 80%以上の認識率となった。一方、テストデータに対する認識率は、5 カテゴリでは 76%となったものの、10, 20 カテゴリでは 50%以下の認識率となった。

7. 移動ロボットの時系列画像を用いた実験

本節では、ロボットの移動とともに得られる視野画像列を用いたカテゴリ分類実験を行う。本実験の対象データセットは見え方が時間とともに変化する時系列画像と

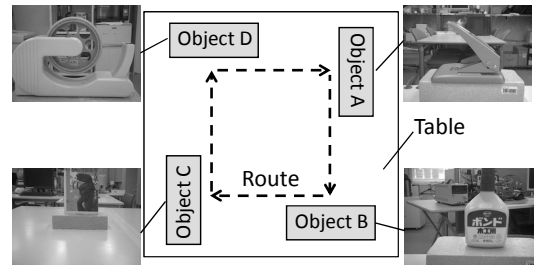


図 8 実験で用いたオブジェクトと走行経路

Fig. 8 Four objects and the robot route used for our experiment.

なり、本研究において動的画像のカテゴリ分類と位置付ける。またデータセットの時系列画像は、独自の実験環境を構築し取得した。以下に実験環境、特徴点の選択結果およびカテゴリ分類の認識率を示す。

7.1 実験環境

本実験は、バンダイ社製のホームロボット「ネットタンサ」を用いた。ロボットの大きさは、高さ 190mm、長さ 160mm、幅 160mm である。搭載されているカメラの仕様は、1/4 インチ CMOS、画像圧縮形式は JPEG、動画像解像度は 320 × 240pixel、フレームレートは 15fps である。動作環境は、1,150 × 1,150mm の机上とし、環境内には 4 つのオブジェクトを配置した。環境内におけるオブジェクトの配置とロボットの走行経路を図 8 に示す。本ロボットの動作環境を机上とし、視覚の高さを考慮し、対象オブジェクトには形状に特徴のある事務用品と本を用いた。対象とするオブジェクトは、図 8 に示すパンチ (Object A)、ボンド (Object B)、本 (Object C)、セロハンテープカッタ (Object D) である。本実験では、本行動における視野画像列をカテゴリ分類の対象データセットとする。また、環境内を 2 周し、1 周目の視野画像列を RUN1、2 周目を RUN2 としてデータセットを構築した。

7.2 OC-SVMs による SIFT 特徴点の選択結果

ロボットの視野画像列における OC-SVMs の SIFT 特徴点の選択結果を図 9 に示す。ロボットからオブジェクトまでの距離の違いによる見え方の変化に対してもオブジェクトを中心とする SIFT 特徴点を選択されている。図 9(d) に示すセロハンテープカッタの画像に着目すると、オブジェクトから離れた位置ではセロハンテープ全体、近い位置ではセロハンテープを回転させる箇所の特徴点を選択されている。また、オブジェクトだけでなくその周辺の SIFT 特徴点も選択された。

7.3 カテゴリ分類結果

RUN1 の視野画像列を用いた実験における ART-2 によるラベル形成の結果を図 10 に示す。縦軸が形成したラベルを、横軸がフレーム数を示す。また、図の上部には

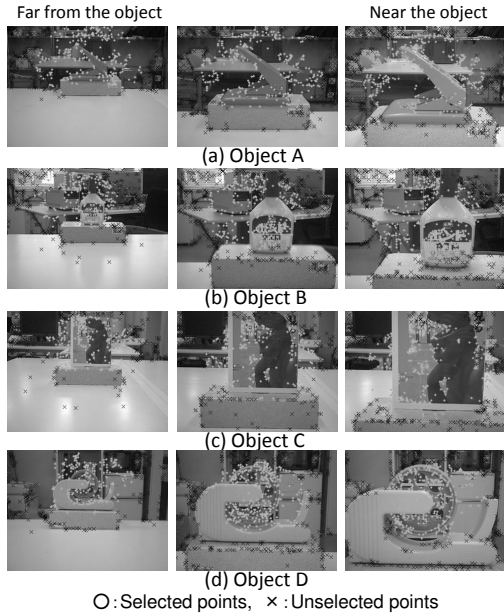


図 9 視野画像列における SIFT 特徴点の選択結果

Fig. 9 Results of selected SIFT feature points of time-series images.

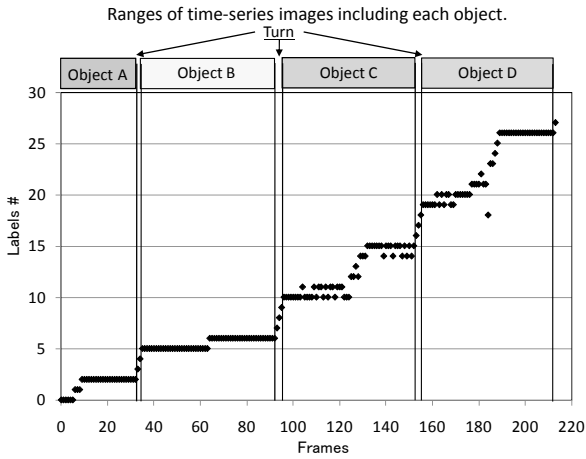


図 10 視野画像列の ART-2 によるラベル形成の結果

Fig. 10 Results of formed labels using ART-2 at time-series images.

表 2 視野画像列における学習とテストの認識率

Table 2 Recognition rates of learning and testing datasets of time-series images.

		Testing Datasets		Mean
		RUN1	RUN2	
Training Datasets	RUN1	98.1%	96.2%	96.7%
	RUN2	97.2%	98.8%	

各オブジェクトが含まれる視野画像列の範囲と、90度旋回の箇所を示す。図 10 より、視野画像列から 27 ラベルが生成された。また、90度旋回する際の画像は、1 画像に 1 ラベルとなり、全体のラベル数を増加させた。各オブジェクトに着目すると、Object A は 3 ラベル、Object B は 2 ラベル、Object C は 6 ラベル、Object D は 8 ラベルとなった。

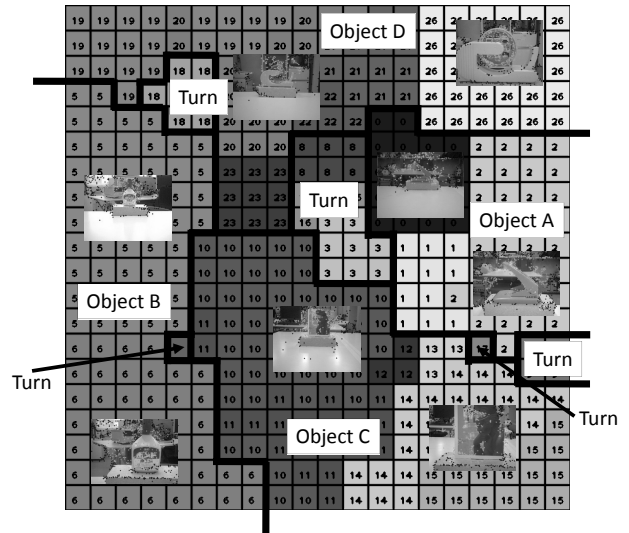


図 11 視野画像列の CPNs によるカテゴリマップの結果

Fig. 11 Mapping result of images on the category map of CPNs used in labels generated by ART-2.

次に、CPNs によるカテゴリマップの結果を図 11 に示す。カテゴリマップ上には、各オブジェクトの画像が写された領域を示す。図に示す結果より、ART-2 で複数のラベルが形成された各オブジェクトの画像は、カテゴリマップの隣接するユニットに写された。また、ラベル 3 や 4 の 90 度旋回する際の画像は、各オブジェクト領域の境界付近に写された。さらに、Caltech-256 の実験と同様に算出した認識率を表 2 に示す。表 2 は、RUN1 と RUN2 を学習とテストそれぞれで用いた際の認識率を示す。また、下線の値は学習データに対する認識率となる。なお、SIFT 特徴量は、オクルージョンが 30% 以上では再現率 50% 以下となることが文献 [14] より述べられている。オブジェクトの欠損が 30% 以上の画像は、背景・その他カテゴリとした。表 2 から、いずれの学習、テストにおいても 90% 以上の高い認識率となった。また、2 つのテスト結果の平均認識率は 96.7% であった。一方、90 度旋回の画像で各オブジェクトのラベルと混同し、誤認識を招いた。

8. 考 察

Caltech-256 およびロボットの視野画像列を用いた実験結果より、OC-SVMs はオブジェクト全体だけではなく、オブジェクト付近の背景や、オブジェクトの一部などの特徴点を選択した。このことから OC-SVMs は、単純なオブジェクトと背景という特徴点の分類ではなく、画像中の特定の情報が集中する箇所、つまり画像を特徴付ける特徴量を選択できているといえる。人間がオブジェクトを分類する際、オブジェクト全体からの判断ではなく、オブジェクトを特徴付ける情報の集中する箇所に注目している。このことから OC-SVMs による SIFT 特徴点の選択は、カテゴリ分類に有効な特徴記述ができ、より高い分類精度に寄与する特徴表現が可能であると考えら

れる。

本実験では、Caltech-256 を用いた静的カテゴリ分類とロボットの視野画像列を用いた動的カテゴリ分類を行った。静的カテゴリ分類では、20 カテゴリに対して学習で 81%、テストで 50% の認識率で分類することができた。Caltech を対象とした教師なしカテゴリ分類法としては、文献 [7] が 26 カテゴリに対して学習で 76.9%、テストで 67.4% の認識率を示している。本手法は、Caltech に対して認識率として十分な結果は得られなかった。しかしながら本手法は、事前にカテゴリ数の設定を必要としない。そのため本来、一般物体認識において困難な課題とされているカテゴリの範囲および種類が曖昧な対象にも適用可能な手法であるといえる。本実験では、ART-2 で 3 ラベル以上の複数のラベルが形成されるカテゴリが 10 カテゴリ観測された。これは、ART-2 が入力データの系列変化を積極的に学習するのに対し、Caltech-256 の画像間に系列要素が存在しないためにラベルの系列性が保たれなかったと考えられる。

一方、動的カテゴリ分類では、学習、テストともに 90% 以上の高い認識率を示した。本実験の分類対象である視野画像列は画像間に系列要素が存在し、本手法の特性を活かしたカテゴリ分類ができたといえる。一般物体認識では、1 つのカテゴリには 1 種類のラベルを与えてカテゴリに分類している。しかしながら、ロボットビジョンにおけるカテゴリ分類では、環境内のセンシングに伴う見え方の変化に対して追加的にラベルを与えてカテゴリに分類する必要がある。本手法の ART-2 のラベル生成では、追加的に見え方の変化を学習できていると評価できる。一方、90 度回転などの行動によって環境内の見え方が増えるため、ART-2 のラベル数は増加する。しかしながら、CPNs では、見え方の異なるオブジェクトの画像を隣接するユニットへ写像し、各オブジェクトのカテゴリを形成した。また、CPNs の近傍学習に基づく位相写像特性により、各オブジェクトを特徴付ける画像は、各カテゴリの中心付近に写像され、回転時などの画像は、カテゴリ間の境界付近に写像される。このことから本手法は、カテゴリ分類とともにカテゴリ内の多様性も表現できるといえる。

本研究では、カテゴリ数が未知である実環境でのカテゴリ分類を目標としている。本実験結果から本手法は、ART-2 と、CPNs を用いることでカテゴリマップ上で時間的な関係性の空間的な可視化を行い、カテゴリ数が未知な実環境のカテゴリ分類に適用可能であると考えられる。本手法は、通常のコンピュータビジョン分野におけるカテゴリ分類だけでなく、カテゴリ数が未知で環境内の見え方も多様なロボットビジョンのカテゴリ分類手法として有効であるといえる。

9. まとめ

本研究では、ART-2 の追加学習機能と CPNs の自己

写像特性を組み合わせた教師なしカテゴリ分類、および OC-SVMs を適用した教師なし SIFT 特徴点選択を提案した。本実験から、画像を特徴付ける情報が集中する箇所の SIFT 特徴点を選択し、カテゴリ分類の精度向上に大きく寄与する特徴表現が可能であることが分かった。また本手法は、ART-2 のラベル生成によってオブジェクトの見え方の変化や形状の違いを表現し、CPNs のカテゴリマップによってそれらを隣接するユニットに写像することで、カテゴリ内の多様性も表現できる。今後の課題は、CPNs のカテゴリマップからクラスタ境界を自動抽出し、適切なカテゴリ数を決定することが挙げられる。また、オブジェクトの分類および認識のためのロボットの行動生成を含めたアプローチの検討が挙げられる。

文 献

- [1] 柳井 啓司, "一般物体認識の現状と今後," 情報処理学会論文誌: コンピュータビジョンとイメージメディア, vol. 48 no. SIG16 (CVIM 19), Nov. 2007.
- [2] 中野 馨, "脳をつくる - ロボット作りから生命を考える -," 共立出版株式会社, Aug. 1995.
- [3] Barnard, K., Duygulu, P., Freitas, N. d., Forsyth, D., Blei, D. and Jordan, M.: Matching Words and Pictures, Journal of Machine Learning Research, vol. 3, pp. 1107-1135, 2003.
- [4] 鈴木 浩二, 松川 徹, 栗田 多喜夫, "サポートベクターマシンを用いた Bag-of-Features における局所特徴の初期特徴選択," 電子情報通信学会 PRMU, 2009.
- [5] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, and W.T. Freeman, "Discovering Objects and their Localization in Images," IEEE International Conference on Computer Vision, pp. 370-377, 2005.
- [6] L. Zhu, Y. Chen, and A. Yuille, "Unsupervised Learning of Probabilistic Grammar - Markov Models for Object Categories," IEEE Trans. PAMI vol. 31, no. 1, Jan. 2009.
- [7] Y. Chen, L. Zhu, A. Yuille, and H. Zhang, "Unsupervised Learning of Probabilistic Object Models (POMs) for Object Classification, Segmentation, and Recognition Using Knowledge Propagation," IEEE Trans. PAMI vol. 31, no. 10, Oct. 2009.
- [8] 中村 友昭, 長井 隆行, 岩橋 直人, "ロボットによる物体のマルチモーダルカテゴリゼーション," 電子情報通信学会論文誌, D vol. J91-D, no. 10, pp. 2507-2518, 2008.
- [9] 寺島 幹彦, 白谷 文行, 山本 公明, "自己組織化特徴マップ上のデータ密度ヒストグラムを用いた教師なしクラスタ分類法," 電子情報通信学会論文誌, D-II vol. J79-D-II, no. 7, pp. 1280-1290, Jul. 1996.
- [10] T. Kohonen, "Self-Organizing Maps," Springer Series in Information Sciences, 1995.
- [11] BayoNet (K2 アルゴリズム)
<http://www.msi.co.jp/BAYONET/files/manual.pdf>
- [12] G.A. Carpenter and S. Grossberg, "ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns," Applied Optics, vol. 26, pp. 4919-4930, 1987.
- [13] R. Hetch-Nielsen, "Counterpropagation networks," Proc. of IEEE First Int'l. Conference on Neural Networks, 1987.
- [14] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," IEEE Trans. PAMI, vol. 27, no. 10, Oct. 2005.