

# 適応的学習による自己組織化マップの教師なしカテゴリ分類

## Unsupervised Category Classification of Self-Organizing Maps Using Adaptive Learning

塚田 真敬  
Masahiro Tsukada

間所 洋和  
Hirokazu Madokoro

佐藤 和人  
Kazuhiro Sato

秋田県立大学  
Akita Prefectural University  
m11a013@akita-pu.ac.jp

**Abstract:** 本論文では、画像データの教師なしカテゴリ分類法として、ART2 ( Adaptive Resonance Theory ) の追加学習機能と CPN ( Counter Propagation Networks ) の自己写像特性を組み合わせたネットワークを提案する。本手法では、SIFT ( Scale-Invariant Feature Transform ) の各特徴点における 128 次元の SIFT 記述子ベクトルから、SOM を用いてコードブックを生成し、そのヒストグラムを特徴量として、ART2 による教師なしラベル形成、CPN によるラベリング、及びカテゴリ間の空間的関係の可視化を行い、教師なしカテゴリ分類を可能とする。

### 1 はじめに

最新の脳科学の知見によると、人間は未来の出来事を予測するために記憶を蓄積し、整理・編集していることが示されている。この事から人間は、世界をカテゴリごとに分類することで未来予測を可能としており、カテゴリ分類は、経験に基づく知的な振る舞いの基盤となっているといえる。また、ロボットビジョンの研究において、ロボットが人間の感覚に則した視覚を獲得するためには、視覚と同時に知識を使わなければならないと述べられている。その際、ロボットビジョンに必要なとされる能力が、カテゴリ分類であるといえる。この事が実現できると、ロボット自身が自分の置かれている環境を視覚情報によって認識し、安全に移動したり、物体を自由に操作するためのビジョンを獲得できるのではないかと考えられている。視覚情報処理分野では、画像中の未知の物体をそれが属すべきカテゴリに分類する一般物体認識が、知識システム獲得のための技術として実現が期待されている [1]。また、画像中に含まれている物体やシーンの認識をするだけでなく、人間が行う「想像」や「記憶」のように画像から予測される様々な可能性について、システムがコンテキストを理解し、画像に写っている世界の状態そのものの記述を行うことが望ましい。こうした事が実現できて、初めて画像の意味的な認識・理解が実現できたといえるのではないかと考えられている。

最近の一般物体認識の研究では、あらかじめ分類カテゴリを決めて、それに対応する学習画像を手で集める教師ありカテゴリ分類法とは異なり、分類カテゴリ自体が未知な問題に対して自動的にクラスを探し出し、画像をカテゴリ分類させる教師なしカテゴリ分類法が提案されている。Sivic らは教師なしカテゴリ分類として、Bag-of-keypoints を用いて、大量の画像に対して文書分類手法の pLSA ( probabilistic Latent Semantic Analysis ) を適用することによって、自動的に画像のカテゴリ分類を行う手法を提案している [2]。Zhu らは、PCFG ( Probabilistic Context-Free Grammars ) と MRF ( Markov Random Fields ) を組み合わせた生成的モデルの PGMM ( Probabilistic Grammar-Markov Models ) を用いて、教師なしカテゴリ分類のために、物体カテゴリのモデル化を教師なしで行っている [3]。中村らは、人間の感覚に即した物体のカテゴリ分類として、ロボットの身体性を利用し、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 [4]。このように近年、視覚情報処理において、知識システム獲得のための技術として、教師なしカテゴリ分類の研究が非常に重要視されている。しかし、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条件があるため、カテゴリ数が未知である実環境でのカテゴリ分類には問題点となる。

そこで本研究では、分類カテゴリ数が未知な問題にも適応できる教師なしカテゴリ分類法として、Grossberg らによって提案された ART2 ( Adaptive Resonance Theory ) [5] の追加学習機能と Nilsen によって提案された CPN ( Counter Propagation Networks ) [6] の自己写像特性を組み合わせたネットワークを提案する。ART2 は、安定性と可塑性を保ちながら適応的にカテゴリを形成する追加学習型の自己組織化ニューラルネットワークである。ART2 を用いることにより、事前にカテゴリ数の知識を必要としない教師なしカテゴリ分類が可能となる。また、ロボットビジョンから得られる画像のカテゴリ分類など、実環境での視覚情報は時間の経過と共に変化し、カテゴリ数も未知であるため、時間軸に沿った学習を進めることができる ART2 の適用は、時系列変化を扱う上でも有用であると考えられる。CPN は、SOM ( Self-Organizing Maps ) の Kohonen 層の上部に Grossberg 層を追加した 3 層で構成され、写像とラベリングを同時に行う教師ありニューラルネットワークである。本手法の特徴は、これらをお互い組み合わせることにより、ART2 では不明瞭となっているカテゴリ間の空間的な関係を CPN により可視化するとともに、ART2 により教師なし学習として CPN のラベル付けを自動で行うことができる点にある。また、SOM は入力特徴量となるコードブック作成におけるクラスタリングにも用いる。画像の特徴量記述には、画像の回転・スケール変化・照明変化などにも頑健な局所特徴量を記述する SIFT ( Scale-Invariant Feature Transform ) 記述子を用いる。

本研究では、大きく 2 つの実験を行うことによって本手法の有効性を示す。1 つ目の実験は、カテゴリ分類研究の評価実験として一般的に用いられている Caltech 画像セットを用いたカテゴリ分類実験を行う。Caltech-256 を用いた評価実験では、ART2 の分類粒度を決定するピジランスパラメータを変更することにより、冗長なカテゴリ変化を抑制しつつ、CPN のラベリングにより各カテゴリの空間

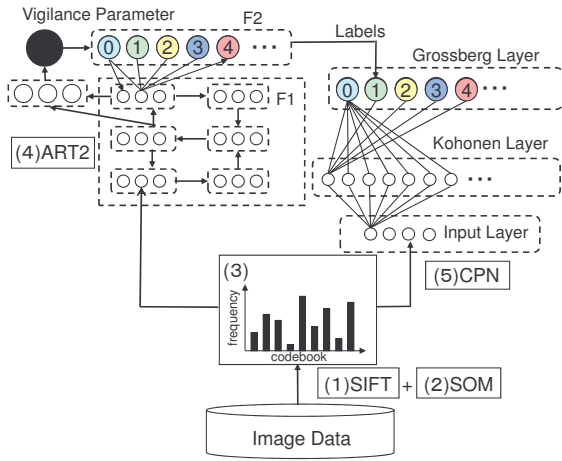


Fig. 1: Network architecture of the proposal method.

関係を写像できるという本手法の特徴を示す。2つ目の実験は、ロボットビジョンから得られる時系列画像を用いたカテゴリ分類実験を行う。ロボットが視覚情報をもとに自律行動を行うためには、視覚から得られる画像情報のカテゴリ分類が必要不可欠である。本手法では、時系列に沿った追加学習機能を持つ ART2 を用いていることから、ロボットの移動とともに得られる画像情報のカテゴリ分類が可能であるといえる。そこで本実験では、ロボット走行環境内にオブジェクトを置き、ロボットの移動とともに見え方が変化するオブジェクトを、同カテゴリとして分類することで、本手法の有効性を示す。

以下、2では提案する教師なしカテゴリ分類法について詳しく説明する。Caltech-256、ロボットビジョンによる実験方法、結果を3,4で示し、5でその考察について述べる。最後に6で結論と今後の課題について述べる。

## 2 教師なしカテゴリ分類法

本節では提案する教師なしカテゴリ分類法について述べる。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 Fig.1 に示す。提案手法の処理手順は Fig.1 のように、(1) SIFT による特徴点抽出および SIFT 記述子ベクトルの算出、(2) 全学習画像の全 SIFT 記述子ベクトルの SOM を用いたクラスタリングによるコードブックの作成、(3) コードブックに基づいて各画像について SIFT 記述子ベクトルのヒストグラムを作成、(4) ART2 による教師なしラベル形成、(5) CPN のラベリングによるカテゴリマップの作成 (カテゴリ分類)、となる。提案手法の前処理にあたる (1)~(3) は Bag-of-keypoints 表現をベースとしている (1) の SIFT 記述子は、現在、一般物体認識において、局所特徴量の記述方法として最もよく用いられている。SIFT の処理は、特徴点の検出 (detection) と特徴量の記述 (description) の2段階からなるアルゴリズムである。また本手法では (2) のクラスタリング手法として SOM を用いる。k-means 法ではクラスタ中心以外のデータ情報が考慮されないのに対し、SOM では位相情報を保存しながらの学習を行うことができるため、誤認識を最小に抑えることができる。更に、(4)、(5) のネットワークシステムにより、大量の画像に自動でラベルを与えてカテゴリに分類する、教師なしカテゴリ分類を可能とする。以下に、各アルゴリズムの詳細を記す。

### 2.1 SOM によるコードブックの作成

本手法では、一般的な Bag-of-keypoints とは違い、k-means 法によるクラスタリングを SOM で行う。SOM のネットワークは、入力データと類似したデータにも反応するような領域を作り出し、入力データの持つ特徴を判別できるように、ネットワークの内部状態を変更しながら学習する。よって、SOM では、学習時に提示されなかったベクトルでも、正確にクラスタリングされ、クラスタ数や初期クラスタの中心値が未知の場合でも誤認識を最小とするクラスタ抽出を行うことができる。なお、SOM の学習アルゴリズムは、2.3 節にて述べる CPN の Grossberg 層を除いた、入力層と Kohonen 層との間の処理となる [7]。

### 2.2 ART2 によるカテゴリのラベル付け

ART には様々なバージョンが存在する [8]。その中で、本手法では、連続値入力可能な ART2 [5] を用いる。ART2 の学習アルゴリズムを以下に記す。

- (1) トップダウン結合荷重  $Z_{ji}$ 、ボトムアップ結合荷重  $Z_{ij}$ 、F1 層サブレイヤの出力  $p_i$ 、 $q_i$ 、 $u_i$  を次式により初期化する。

$$Z_{ji}(0) = 0, \quad Z_{ij}(0) = \frac{1}{(1-d)\sqrt{M}}, \quad (1)$$

$$p_i(0) = q_i(0) = u_i(0) = v_i(0) = w_i(0) = x_i(0) = 0.0 \quad (2)$$

- (2) 入力  $I_i$  を提示し、次式より各サブレイヤを伝播させる。

$$w_i(t) = I_i(t) + au_i(t-1), \quad (3)$$

$$x_i(t) = \frac{w_i(t)}{e + \|w\|}, \quad (4)$$

$$v_i(t) = f(x_i(t)) + bf(q_i(t-1)), \quad (5)$$

$$u_i(t) = \frac{v_i(t)}{e + \|v\|}, \quad (6)$$

$$p_i(t) = \begin{cases} u_i(t) & (\text{inactive}) \\ u_i(t) + dZ_{ji}(t) & (\text{active}) \end{cases} \quad (7)$$

$$q_i(t) = \frac{p_i(t)}{e + \|p\|}, \quad (8)$$

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < \theta \\ x & \text{if } x \geq \theta \end{cases} \quad (9)$$

- (3) 最大活性化ユニット  $T_j$  を探索する。

$$T_J(t) = \max_j \left( \sum_j p_i(t) Z_{ij}(t) \right) \quad (10)$$

- (4) トップダウン結合荷重  $Z_{ji}$ 、ボトムアップ結合荷重  $Z_{ij}$  を更新する。

$$\frac{d}{dt} Z_{ji}(t) = d[p_i(t) - Z_{ji}(t)], \quad (11)$$

$$\frac{d}{dt} Z_{ij}(t) = d[p_i(t) - Z_{ij}(t)] \quad (12)$$

- (5) ビジランス閾値  $\rho$  を用いて、カテゴリに属するかを判定する。

$$\frac{\rho}{e + \|r\|} > 1, r_i(t) = \frac{u_i(t) + cp_i(t)}{e + \|u\| + \|cp\|} \quad (13)$$

判定条件が成立する場合は、選択されたユニットをリセットして(3)へ戻り再探索する。不成立の場合は、F1層内の変化率が小さくなるまで(2)と(4)を繰り返す。なお、 $a$ と $b$ は $u$ 層から $w$ 層及び $q$ 層から $v$ 層のフィードバックループにかかる係数である。 $c$ は $p$ 層から $r$ 層への伝搬係数、 $d$ は学習率係数であり、両者間には $cd/(1-d) \leq 1$ の制約が与えられている。 $\theta$ は $v$ 層におけるノイズの除去を制御するパラメータである。なお、本手法では $\theta$ を0.1に設定した。

### 2.3 CPNによるカテゴリマップの生成

CPNはパターンマッピングを行う[6]。すなわちパターンの全集合において一つのパターンを別なパターンにマッピングする。パターンが提示されると、学習後のネットワークは結合荷重を使って、そのパターンを特定のグループに分類する。本手法では、通常、実験者や設計者の経験や主観的な判断による手作業に頼っているKohonen層ユニットに対するラベル付けをART2により自動で行っている。

CPN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。

$$d_{n,m} = \sqrt{\sum_{i=1}^I (x_i(t) - u_{n,m}^i(t))^2} \quad (14)$$

- (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。

$$c = \operatorname{argmin}(d_{n,m}) \quad (15)$$

- (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。

$$u_{n,m}^i(t+1) = u_{n,m}^i(t) + \alpha(t)(x_i(t) - u_{n,m}^i(t)) \quad (16)$$

- (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。

$$v_{n,m}^j(t+1) = v_{n,m}^j(t) + \beta(t)(t_j(t) - v_{n,m}^j(t)) \quad (17)$$

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。CPN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。なお、本手法では、学習率係数  $\alpha(t)$  および  $\beta(t)$  を0.5、学習回数は10000回とした。

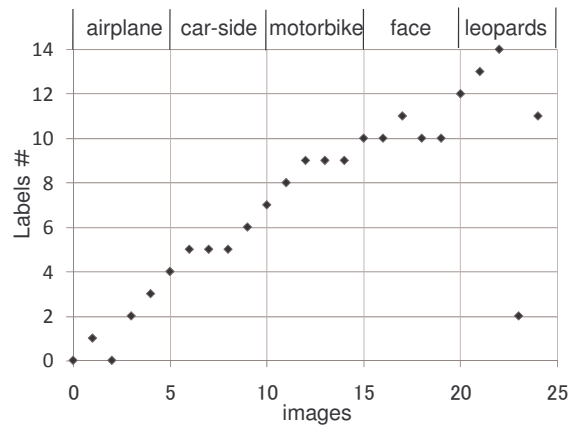


Fig. 2: Results of ART2.

### 3 Caltech-256を用いた実験

近年のカテゴリ分類の研究では、カリフォルニア工科大学のCaltech画像セットが評価画像データのデファクトスタンダードとなっている[1]。そこで、本研究では、最新のCaltech-256を用いてカテゴリ分類実験を行う。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(他手法[2],[3]でも使用)を用いた、5カテゴリ、10カテゴリ、20カテゴリを用いた静止画像の分類実験から、本手法の性能評価を行った。なお、ART2のビジランスパラメータは0.930とした。

5カテゴリにおけるART2の分類結果をFig.2に示す。なお、縦軸はラベル、横軸はデータ数を示し、グラフ上部にはグランドトゥルースを示す。ART2の分類結果から、各画像へ独立したラベルを形成し、5カテゴリを分類できた。leopardsカテゴリでは、airplane、faceカテゴリとラベルが混同した。また、CPNのカテゴリマップをFig.3に示す。CPNのカテゴリマップから、ART2のラベルを基に各カテゴリの画像を近い位置のユニットへ写像することで、各カテゴリを形成することができた。また、CPNでは、ART2でラベルが混同したleopardsカテゴリの画像を、近い位置のユニットへ写像することでカテゴリを形成することができた。CPNのカテゴリ分類結果のグランドトゥルースとの一致率は、25枚中1枚で誤分類があったため96.0%となった。また、10カテゴリの分類では、50枚中14枚で誤分類があったため72.0%、20カテゴリの分類では、100中47枚で誤分類があったため53.0%であった。ART2のラベル数は、10カテゴリで29ラベル、20カテゴリで41ラベルとなり、5カテゴリと比較すると大幅に増加した。

### 4 移動ロボットを用いた実験

ロボットビジョンから得られる走行シーンの時系列画像を用いたカテゴリ分類実験を行った。本実験では、ロボットの走行環境内に置かれたオブジェクトのカテゴリ分類実験から、本手法の性能評価を行った。

#### 4.1 実験環境

本実験では、Fig.4に示す、バンダイ社製のホームロボット「ネットタンサー」を用いた。ネットタンサーの概要は、全高約190mm、全長約160mm、全幅約160mmである。搭載されているカメラは、1/4インチCMOSの30万画素、画像圧縮形式はJPEG、カメラ解像度は320 × 240ppi、フ



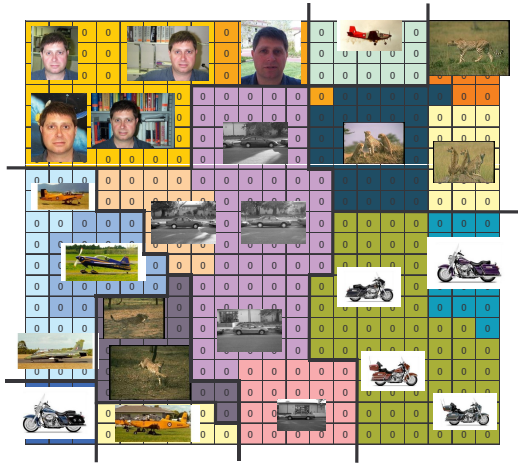


Fig. 3: Results of CPN.

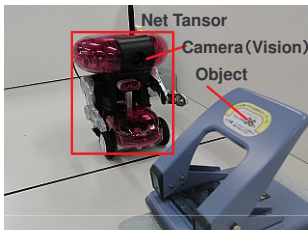


Fig. 4: The Robot which Used for an Experiments(Net Tensor).

フレームレートは 15fps である。分類対象データは、処理時間を考慮して毎秒 3 フレームを取り込む。ロボットの走行は、高さ 300mm の白い壁で囲んだ 1150 × 1150mm の環境内で行う。走行環境内には、Fig.5 に示すパンチ (objectA)、セロハンテープ (objectB)、本 (objectC)、ボンド (objectD) の 4 つのオブジェクトを配置した。ネットタンサーの視覚の高さを考慮し、実環境での動作範囲を机上と想定したため、対象オブジェクトとして形状に特徴のある事務用品を用いた。本実験では、ロボットの走行とともに見え方の変化する 4 つのオブジェクトに背景画像を加えた 5 つのカテゴリ进行分类し、本手法の性能評価を行う。実験は、2 オブジェクト、3 オブジェクト、4 オブジェクトそれぞれで行い、その際のロボットの走行は Fig.6 の (a)、(b)、(c) に示すオブジェクトの配置と走行経路で行った。4 オブジェクトでは、走行環境を 2 周させて、2 周目の時系列画像をテストデータとしてテストの分類も行った。なお、本実験における ART2 のビジランスパラメー



Fig. 5: Objects and Background.

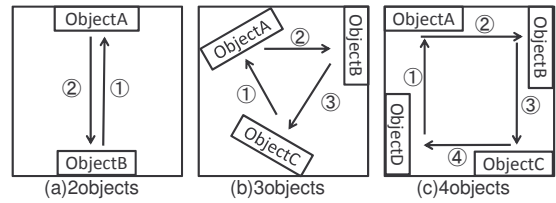


Fig. 6: The Environment of Experiments and The Run Course of Robot.

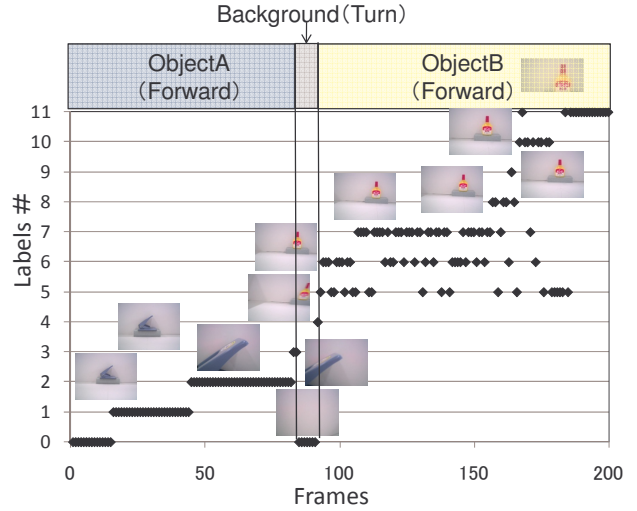


Fig. 7: Results of ART2 at 2objects.

タは 0.850 とした。

#### 4.2 2 オブジェクトおよび 3 オブジェクト

本実験では、Fig.6 (a)、(b) のようにオブジェクトを配置して、ロボットが走行した時に得られる時系列画像のカテゴリ分類を行った。2 オブジェクトにおける ART2 の分類結果を Fig.7、CPN のカテゴリマップを Fig.8 に示す。Fig.7 の結果より、Fig.6 (a) の走行における時系列画像は 12 ラベルに分類された。分類された各ラベルと画像の関係は、ロボットがターンする際に見える背景画像が # 0 の 1 ラベル、objectA が写った画像が # 0~3 の 4 ラベル、objectB が写った画像が # 4~11 の 8 ラベルに分類された。よって、# 0 にラベル付けされた objectA の画像と背景画像は、ラベルが混同した。Fig.8 の CPN のカテゴリマップでは、ART2 のラベルを基に時系列画像を各オブジェクトで近い位置のユニットへ写像することで、各カテゴリを形成することができた。背景画像は、ART2 でラベルが混同したため、CPN では位置関係が離散的となって写像された。CPN のカテゴリ分類結果のグランドトゥールースとの一致率は、200 枚中 7 枚で誤分類があったため、96.5 % となった。

3 オブジェクトにおける ART2 の分類結果を Fig.9、CPN のカテゴリマップを Fig.10 に示す。Fig.9 の結果より、Fig.6 (b) の走行における時系列画像は 13 ラベルに分類された。分類された各ラベルと画像の関係は、ロボットが 120 度ターンする際に見える背景画像が # 2 の 1 ラベル、objectA が # 0 と # 1 の 2 ラベル、objectB が # 3~6 の 4 ラベル、objectC が # 7~12 の 6 ラベルに分類された。2 オブジェクト実験における各オブジェクトのラベル数を比較すると、objectA、B ともにラベル数は減少した。Fig.10 の CPN のカテゴリマップでは、ART2 のラベルを基に時系列画像を objectA、objectB、objectC、背景画像で近い位置のユニットへ写像し、各カテゴリを形成することができた。CPN

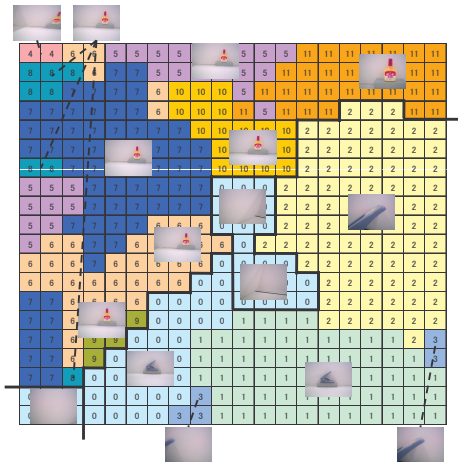


Fig. 8: Results of CPN at 2 objects.

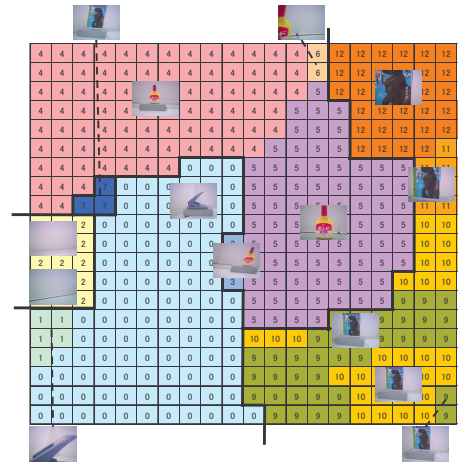


Fig. 10: Results of CPN at 3 objects.

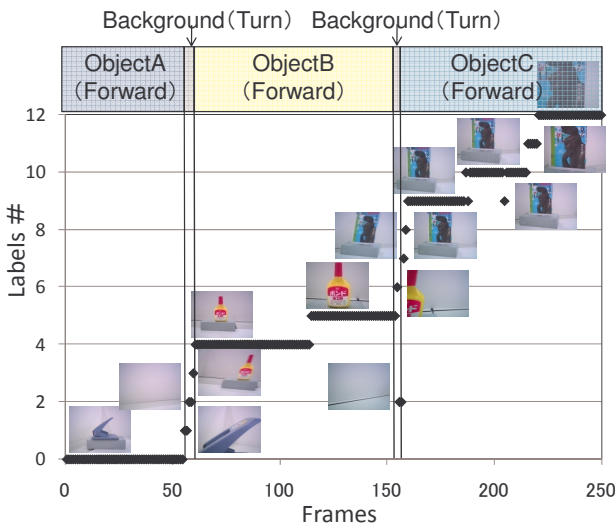


Fig. 9: Results of ART2 at 3 objects.

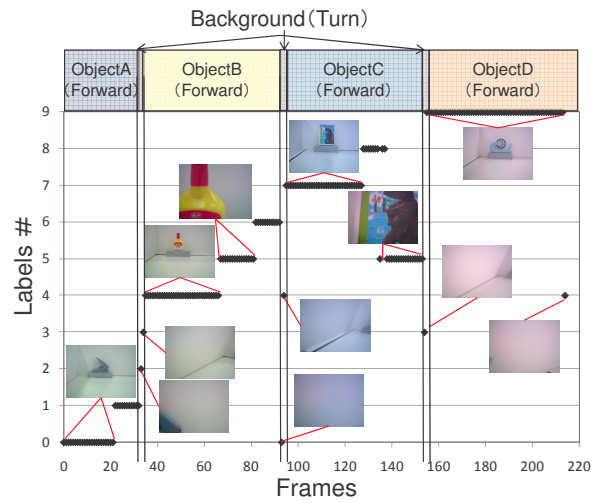


Fig. 11: Results of ART2 at 4 objects.

のカテゴリ分類結果のグランドトゥールースとの一致率は、250 枚中全ての画像が objectA, objectB, objectC, 背景画像と分類できたため、100 %となった。

### 4.3 4 オブジェクト

本実験では、Fig.6 (c) のように objectA, B, C, D を配置して、ロボットが走行した時に得られる時系列画像のカテゴリ分類を行った。また、本実験では走行環境を2周した。2周目の時系列画像をテストデータとして、テストのカテゴリ分類も行う。ART2 の分類結果を Fig.11, CPN のカテゴリマップを Fig.12, テストの分類結果を Fig.13 に示す。Fig.11 の結果より、Fig.6 (c) の走行における時系列画像は 10 ラベルに分類された。分類された各ラベルと画像の関係は、ロボットが 90 度ターンする際に見える背景画像は # 3 の 1 ラベルに分類された。また、objectA は # 0, # 1, # 2 の 3 ラベル、objectB は # 4, # 5, # 6 の 3 ラベル、objectC は # 7, # 8 の 2 ラベル、objectD は # 9 の 1 ラベルに分類された。近づきすぎて objectC の一部しか写らなかった画像、90 度ターンの際に背景が写った 94, 95, 215 フレームの画像は、ラベルが混同した。各オブジェクトのラベル数を 2 オブジェクト、3 オブジェクト実験と比較すると、4 オブジェクトでは全てのオブジェ

クトでラベル数が減少した。このことから、4 オブジェクトとなっても、各オブジェクトの見え方の変化を同じカテゴリに分類できたといえる。しかし、objectB と objectC に近づいた画像は、同カテゴリとして分類された。Fig.12 の CPN のカテゴリマップでは、ART2 のラベルを基に時系列画像を各オブジェクトで近い位置のユニットへ写像し、各カテゴリを形成することができた。ART2 の分類でラベルが混同し、同カテゴリとして分類された objectB と objectC に近づいた画像は、近い位置のユニットへ写像された。CPN のカテゴリ分類結果のグランドトゥールースとの一致率は、215 枚中 20 枚で誤分類があったため、90.7 %となった。テストの分類結果を示した Fig.13 では、右部にグランドトゥールースラベル、上部に時系列画像のグランドトゥールースカテゴリを示す。よって、両方の重なる点線の範囲から外れた画像は、誤分類となる。誤分類となった画像を丸で示した、215 枚中 17 枚で誤分類があったため、一致率は 92.3 %となった。

## 5 考察

本実験では、Caltech-256 の静止画像と移動ロボットの時系列画像を用いたカテゴリ分類を行った。5 カテゴリの実験では、ART2 での 2 画像のラベルの混同を、CPN では 1 画像に減少した。これは、ART2 の追加学習機能と CPN

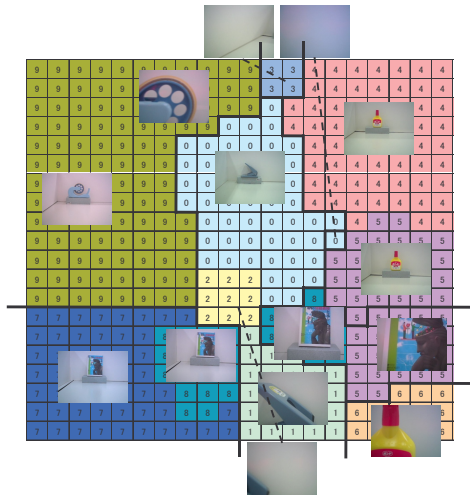


Fig. 12: Results of CPN at 4objects.

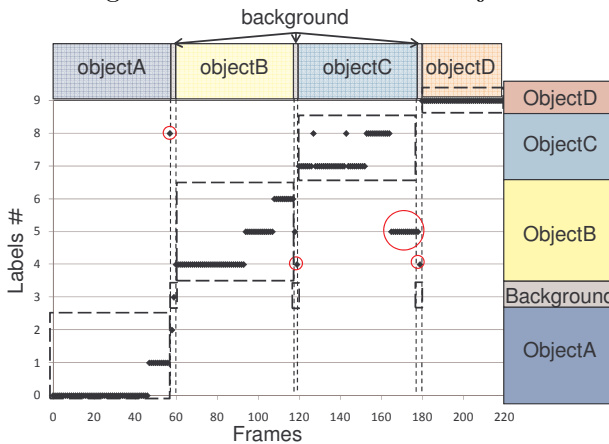


Fig. 13: Test Results of CPN at 4objects.

の高い写像能力を組み合わせた本手法の特徴であると考えられる。10カテゴリ、20カテゴリの結果からは、ART2のラベル数が増加するとCPNのカテゴリ分類の一致率が低下することがわかった。これは、ART2のラベル付けが、1画像に1ラベルとなることが多くなったことで、CPNのカテゴリ分類性能の低下に影響を与えたと考えられる。また、ART2では、入力データの系列変化を積極的に学習する。一方、Caltech-256のデータセットは、画像間に系列要素がなく、カテゴリ内でもオブジェクトの形状や背景が大きく異なる。このため、当該データセットでは、ART2の本来の追加学習機能が発揮されず、ラベルの系列性が保たれなかったと考えられる。

時系列画像を用いた実験では、ART2の特性を発揮できたため、全ての実験で高い一致率を示した。また、オブジェクト数を増やしていくことで、オブジェクトごとのラベル数は抑制され、見え方の変化に対しても頑健にラベル付けができたといえる。これは、複数のオブジェクトを用いてコードブックを生成したことにより、頑健な入力特徴量となったと考えられる。本実験では、ART2で各オブジェクトの見え方の変化を複数のラベルに分けても、CPNのカテゴリマップで各オブジェクトのカテゴリを形成できた。このことから、ART2のラベル数は、カテゴリ内の多様性を表現できるという本手法の特徴であると考えられる。よって、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ため、実環境問題に適した分類手法

であるといえる。また、4オブジェクトで行ったテストの分類では、学習と同程度の一致率を示したため、本手法はテストデータに対しても分類可能であるといえる。しかし、本実験では、オブジェクトに対して前進するという単純な走行だったため、動作とともに対象オブジェクトが拡大するという見え方の変化しか得られていない。環境内に存在するオブジェクトを分類し認識するためには、様々な見え方を得る必要があるため、今後はロボットの行動生成を含めたアプローチを検討する必要があると考えている。

## 6 まとめ

本研究では、画像データの教師なしカテゴリ分類法として、ART2の追加学習機能とCPNの自己写像特性を組み合わせたネットワークを提案し、Caltech-256を用いた実験と移動ロボットの時系列画像を用いた実験から、本手法の特徴と有効性を示した。ロボットビジョンの走行シーンにおける時系列画像を対象とした実験では、5カテゴリ(4オブジェクト+背景)を分類することができたため、対象カテゴリ分類として本手法が有効であることを示すことができた。本研究から、ART2で他カテゴリとラベルが混同した画像を、CPNのカテゴリマップでは同カテゴリと近い位置のユニットに写像することで、カテゴリを形成できることがわかった。また、カテゴリ分類とともにART2のラベルがカテゴリ内の多様性を表現できることがわかった。更に、本手法はテストデータに対しても分類可能であることがわかった。今後の課題としては、SIFTによる特徴量の記述領域を絞り込み、背景変化に対しても頑健な入力特徴量の生成と、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。

## 参考文献

- [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] Josef Sivic, Bryan C. Russell, Alexei A. Efros, Andrew Zisserman, William T. Freeman, "Discovering Objects and their Localization in Images", IEEE International Conference on Computer Vision, pp370-377, 2005.
- [3] Long(Leo) Zhu, Yuanhao Chen, and Alan Yuille, "Unsupervised Learning of Probabilistic Grammar-Markov Models for Object Categories" IEEE, runs. PAMI VOL.31, NO.1, Jan 2009.
- [4] 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.
- [5] Carpenter, G.A. and Grossberg, S., "ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns", Applied Optics, Vol.26, pp4919-4930, 1987.
- [6] Hetch-Nielsen, R. 1987. Counterpropagation networks, Proc. of IEEE First Int'l. Conference on Neural Networks. 1987. :19-32.
- [7] T.Kohonen, "Self-Organizing Maps, Springer Series in Information Sciences", 1995.
- [8] Carpenter, G.A. and Grossberg, S., "Pattern Recognition by Self-Organizing Neural Networks", The MIT Press, 1991.