

## 適応的学習による自己組織化マップの自動カテゴリ分類

塚田 真敬<sup>†</sup> 間所 洋和<sup>†</sup> 佐藤 和人<sup>†</sup>

<sup>†</sup> 秋田県立大学システム科学技術学部 〒015-0055 秋田県由利本荘市土谷字海老ノ口 84-4

E-mail: †{B07A059,madokoro,ksato}@akita-pu.ac.jp

あらまし 本論文では, ART-2(Adaptive Resonance Theory) の追加学習機能, 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用いた, 画像の自動カテゴリ分類手法を提案する. 本手法では, SIFT による 128 次元の特徴量から得られたコードブックに基づいた, SIFT 記述子ベクトルのヒストグラムを ART, CPN に与えることにより, ART による自動ラベリング, CPN によるカテゴリ間の空間的な関係の可視化を行い, unsupervised による自動カテゴリ分類を可能とする. Caltech 画像セットを用いた評価実験では, ART の分類粒度を決定するビジランスパラメータを変更することにより, 冗長なカテゴリを抑制しつつ, CPN のカテゴリマップにより, 自動的にカテゴリ空間を形成できることを示す.

キーワード ART-2, CPN, SOM, SIFT, 自動カテゴリ分類

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

Masahiro TSUKADA<sup>†</sup>, Hirokazu MADOKORO<sup>†</sup>, and Kazuhito SATO<sup>†</sup>

<sup>†</sup> Faculty of Systems Science and Technology, Akita Prefectural University 84-4 Aza Ebinokuchi Tsuchiya, Yurihonjo City, 015-0055 Japan

E-mail: †{B07A059,madokoro,ksato}@akita-pu.ac.jp

**Abstract** This paper presents a method for automatic category classification using a network system that combine additional learning function of ART-2(Adaptive Resonance Theory) networks, self-representation characteristic of CPN(Counter Propagation Networks), and local feature descriptors by SIFT(Scale-Invariant Feature Transform). In our method, we are possible unsupervised automatic category classification. Our proposed method consists of three important elements: 1) automatic category labeling by ART networks; 2) visiblize spatial relations between categories by CPN ;and 3)produce our original histograms basing on vector quantizing SIFT descriptors. Experiment results that used Caltech-256 Object Category Dataset show that our method can successfully build categories from an unsupervised analysis of images.

**Key words** ART-2, CPN, SOM, SIFT, Automatic category classification

### 1. はじめに

画像中の未知の物体をそれが属するべきカテゴリに分類する一般物体認識は, 視覚情報処理における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, 含まれ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さまざまな可能性について, 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することが望ましい. こうしたことが実現できて, 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. 現在, 一般物

体認識研究において最新のカテゴリ分類の手法は, 局所特徴量を用いた Bag-of-keypoints 表現による画像分類手法である. Bag-of-keypoints による特徴表現は, 学習画像に対する, (1) 特徴点抽出 (100 個以上/画像), (2) SIFT 記述子ベクトルの計算, (3) 全学習画像の全 SIFT 記述子ベクトルの k-means を用いたクラタリング (k は 100 ~ 1000 程度) によるコードブックの作成, (4) 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, の手順によって作成される. これを判別モデルにおいて最も高い性能を示している SVM(Support Vector Machine) [2] に適用し, きわめて良い性能を示している. また, 同じく局所特徴量を用いた手法とし

て、GC(Generalized Correlograms) による画像表現手法がある [3]。この画像表現手法では、オブジェクトは GC の分布配置によって表現され、各々の GC は局所部分の情報と、それらの空間関係をコード化している。カテゴリ分類の手法には計算効率に優れた Boosting [4] が用いられており、Caltech 画像セットに対して高い分類精度を示している。

しかし、最近では、これらのあらかじめカテゴリを決めて、それに対応する学習画像を手手で集める supervised な手法とは異なり、カテゴリ自体が未知な問題に対して、画像を自動的にカテゴリ分類させる unsupervised な手法が提案されている。Sivic らは大量の画像から自動的にカテゴリを探し出す unsupervised な手法の試みで、認識すべきカテゴリを自動発見するという興味深いアイデアを提案している [5]。これは Bag-of-keypoints approach を用いて、大量の画像に対して文書分類手法の probabilistic Latent Semantic Analysis(pLSA) を適用することによって、自動的に画像のカテゴリを抽出する手法である。また、文献 [6] では、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 unsupervised なカテゴリ分類手法を提案している。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、unsupervised なカテゴリ分類の研究が非常に重要とされている。

そこで、本研究では、unsupervised なカテゴリ分類を実現する手法として、Grossberg らによって提案された ART-2(Adaptive Resonance Theory) [7] と Nilsen によって提案された CPN(Counter Propagation Networks) [8] の自己写像特性を組み合わせたネットワークを適用した、unsupervised なカテゴリ分類システムを提案する。ART 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。ART 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 unsupervised なカテゴリ分類が可能となる。CPN は SOM(Self-Organizing Maps) の Kohonen 層の上に Grossberg 層を追加し、写像とラベル付けを同時に行う、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、ART では不明瞭となっているカテゴリ間の空間的な関係を SOM により可視化するとともに、教師なし学習として SOM のラベル付けを自動で行える点にある。また、SOM は入力特徴量となるコードブックの作成にも用いており、入力特徴量の記述方法としては、画像の回転・スケール変化・照明変化などに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。Caltech-256 を用いた評価実験では、ART によって未知なカテゴリのラベル付けを自動的に検出できることを示し、提案手法による unsupervised なカテゴリ分類の有効性を検討する。

以下、2. では提案するカテゴリ分類法について詳しく説明する。Caltech-256 を用いた評価実験の方法、結果を 3. で示し、4. でその考察について述べる。最後に 5. で結論と今後の課題について述べる。

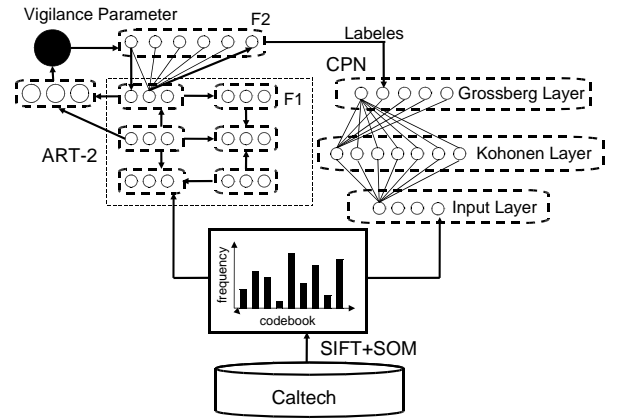


図 1 提案手法のネットワーク構成  
Fig. 1 Network architecture of the proposal method.

## 2. カテゴリ分類法

本節では提案するカテゴリ分類手法について述べる。本研究で提案するカテゴリ分類手法のネットワーク構成を図 1 に示す。提案手法の処理手順は図 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) ART によるカテゴリの自動ラベリング、(5)CPN 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 表現と同様の手順であるが、本手法では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えるため、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、大量の画像から自動的にカテゴリを探し出す unsupervised な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 SIFT による特徴点の検出と特徴量の記述

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は、SIFT 記述子が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 段階からなり、各処理は以下の流れで行われる [9]。

(1) スケールとキーポイント検出 DoG の結果の画像を  $D(u, v, \sigma)$  とすると、DoG 画像は次式で求まる。

$$\begin{aligned} D(u, v, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(u, v) \\ &= L(u, v, k\sigma) - L(u, v, \sigma) \end{aligned} \quad (1)$$

ここで、 $G(x, y, \sigma)$  はガウス関数、 $I(u, v)$  は入力画像、 $L(u, v, \sigma)$  は平滑化画像とする。

DoG 画像の注目画素と、その周りの 26 近傍を比較し、極値であった場合、その画素をキーポイント候補点として検出する。

(2) キーポイントのローカライズ

(2-1) 主曲率によるキーポイントの絞込み。

次に示すようにしきい値処理することで、不要なキーポイ

ントを削除する．

$$\frac{\text{Tr}(\mathbf{H})^2}{\text{Det}(\mathbf{H})} < \frac{(\gamma_{th} + 1)^2}{\gamma_{th}} \quad (2)$$

ここで,  $\text{Tr}(\mathbf{H})$  はヘッセ行列の対角成分の和,  $\text{Det}(\mathbf{H})$  は行列式とする．

DoG の値からキーポイント削除の判別を行う．サブピクセル位置での DoG 出力の絶対値がしきい値より小さい場合キーポイントを削除する．

(2-2) コントラストによるキーポイントの絞込み．

DoG の値からキーポイント削除の判別を行う．サブピクセル位置での DoG 出力の絶対値がしきい値より小さい場合キーポイントを削除する．

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D}{\partial x} \hat{x} \quad (3)$$

ここで,  $D$  は DoG 関数,  $\hat{x}$  はサブピクセル位置とする．

(3) オリエンテーションの算出局所領域における勾配強度  $m(x, y)$  と勾配方向  $\theta(u, v)$  から重み付方向ヒストグラム  $h$  を以下の式により作成する．

$$h_{\theta'} = \sum_x \sum_y w(x, y) \cdot \delta[\theta', \theta(x, y)] \quad (4)$$

$$w(x, y) = G(x, y, \sigma) \cdot m(x, y) \quad (5)$$

(4) 特徴量の記述

特徴量の記述には,  $4 \times 4$  の領域にそれぞれ 8 方向ヒストグラムを作成するため, 128 次元ベクトルの特徴量を持つことになる．この 128 次元の SIFT 特徴量をキーポイントごとに抽出する．

## 2.2 SOM によるコードブックの作成

提案手法では, 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, 2.4 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 [10].

## 2.3 ART ネットワーク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 [11]．その中で, 本研究では, 連続値入力が可能な ART2 [7] 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．

$$Z_{ji}(0) = 0, \quad Z_{ij}(0) = \frac{1}{(1-d)\sqrt{M}}, \quad (6)$$

$$p_i(0) = q_i(0) = u_i(0) = v_i(0) = w_i(0) = x_i(0) = 0.0 \quad (7)$$

(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．

$$w_i(t) = I_i(t) + au_i(t-1), \quad (8)$$

$$x_i(t) = \frac{w_i(t)}{e + \|w\|}, \quad (9)$$

$$v_i(t) = f(x_i(t)) + bf(q_i(t-1)), \quad (10)$$

$$u_i(t) = \frac{v_i(t)}{e + \|v\|}, \quad (11)$$

$$p_i(t) = \begin{cases} u_i(t) & (\text{inactive}) \\ u_i(t) + dZ_{ji}(t) & (\text{active}) \end{cases} \quad (12)$$

$$q_i(t) = \frac{p_i(t)}{e + \|p\|}, \quad (13)$$

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < \theta \\ x & \text{if } x \geq \theta \end{cases} \quad (14)$$

(3) 最大活性化ユニット  $T_j$  を探索する．

$$T_j(t) = \max_j \left( \sum_j p_i(t) Z_{ij}(t) \right) \quad (15)$$

(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．

$$\frac{d}{dt} Z_{ji}(t) = d[p_i(t) - Z_{ji}(t)], \quad (16)$$

$$\frac{d}{dt} Z_{ij}(t) = d[p_i(t) - Z_{ij}(t)] \quad (17)$$

(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．

$$\frac{\rho}{e + \|r\|} > 1, r_i(t) = \frac{u_i(t) + cp_i(t)}{e + \|u\| + \|cp\|} \quad (18)$$

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．

## 2.4 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 [8]．すなわちパターン of 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．そこで結合荷重に関連した目標パターンが出力となる．Kohonen 層では, 教師なし学習を行う競合ユニットをもっている．最上層の Grossberg 層は, Kohonen 層と完全結合して競合的ではない．Grossberg 層は Windrow-Hoff あるいは Grossberg の規則により訓練される．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.

$$d_{n,m} = \sqrt{\sum_{i=1}^I (x_i(t) - u_{n,m}^i(t))^2} \quad (19)$$

(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.

$$c = \operatorname{argmin}(d_{n,m}) \quad (20)$$

(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.

$$u_{n,m}^i(t+1) = u_{n,m}^i(t) + \alpha(t)(x_i(t) - u_{n,m}^i(t)) \quad (21)$$

(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.

$$v_{n,m}^j(t+1) = v_{n,m}^j(t) + \beta(t)(t_j(t) - v_{n,m}^j(t)) \quad (22)$$

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.

### 3. 実験

#### 3.1 画像データセット

近年のカテゴリ分類研究では, 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]. Caltech 画像セットは Caltech-6, Caltech-101, Caltech-256 の 3 種類の画像セットがあり, 本研究では, 最新の Caltech-256 を用いて提案手法の評価を行う. Caltech-6, Caltech-101 と比べ, Caltech-256 はカテゴリが 256 種類に増えただけでなく, Caltech-101 作成時に行った向きや大きさを統一するような操作が行われなかったため, より一般的な画像セットになっているという特徴がある. Caltech-256 は 256 種類のカテゴリと背景カテゴリの計 257 種類から構成されているが, 本実験では 256 種類のカテゴリの中から, 各カテゴリの画像枚数が多く, 他手法 [3], [5] でも用いられている, airplane, car-side, motorbike, face, clutter の 5 種類のカテゴリを用いて本手法の有効性を検討する.

#### 3.2 予備実験

予備実験では本手法の各パラメータとカテゴリ分類性能の関係性を実験より示すとともに, 評価実験で用いるパラメータの検討を行う. 以下に予備実験の結果を示す.

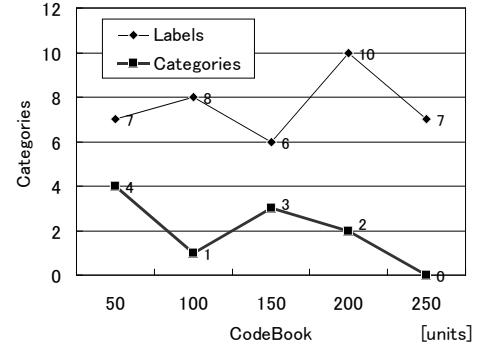


図 2 コードブックのユニット数と ART のラベリング結果  
Fig. 2 The number of codebook units and labeling results by ART.

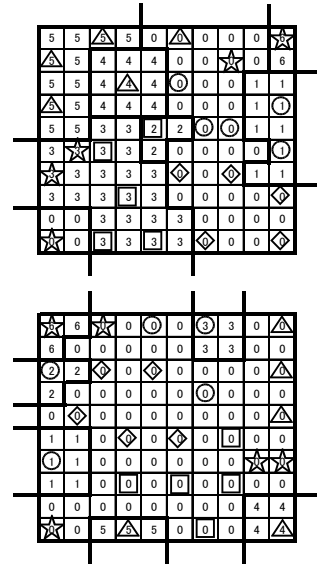


図 3 コードブックのユニット数と CPN のカテゴリマップの比較 (上: ユニット数=50, 下: ユニット数=250)  
Fig. 3 Comparison of units with category maps of CPN (Top: Unit=50, Bottom: Unit=250)

#### 3.2.1 コードブックのユニット数とカテゴリ形成

本手法のカテゴリ分類では入力特徴量として SIFT 特徴量に基づいたコードブックを SOM により生成した. 分類を行う際の特徴量となるコードブックのユニット数の違いがカテゴリ形成に影響を与えるため, ここではユニット数の違いによるカテゴリ形成の比較を行う. ユニット数は 50, 100, 150, 200, 250 で比較を行った. なお, ART のビジランスパラメータ  $\rho$  は 0.950, 学習画像データ数は 5 枚とした. 実験結果を図 2, 3 に示す. 図 2 はコードブックのユニット数の変化と ART のラベリング結果のグラフ, 図 3 はユニット数の変化と CPN のカテゴリマップの比較である. 図 2 のグラフは, Labels が ART でラベリングされたラベルの数を示す, Categories は CPN のカテゴリマップで形成された, カテゴリ数を示す. また, 図 3 のマップに示された番号はラベルを示し, 丸, 四角, 三角, ひし形, 星の記号はそれぞれ, airplane, car-side, motorbike, face, clutter カテゴリを表し, CPN によって写像された位相を示している. はじめに, ART によるラベリングの結果から比較する. 図 2 より, ユニット数が多くなるにつれてラベリングが細分化され, カテゴリが離散的になっていることがわかる.

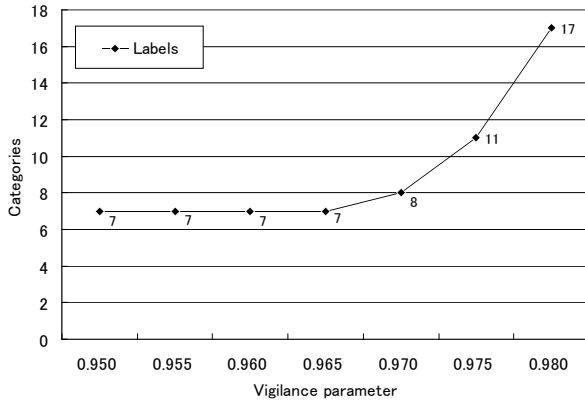


図 4 ビジランスパラメータと ART のラベリング結果  
Fig. 4 The setting of vigilance parameter and labeling results by ART.

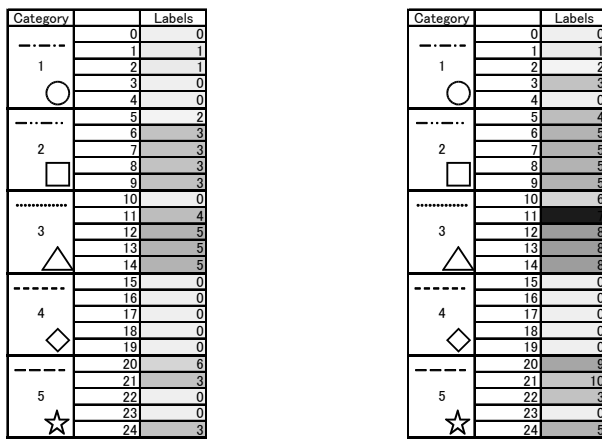


図 5 ART のラベリング結果の比較 (左: $\rho=0.950$ , 右: $\rho=0.975$ )  
Fig. 5 Comparison of labeling results by ART(left: $\rho=0.950$ , right: $\rho=0.975$ ).

また、最も違いの現れたユニット数 50 と 250 の CPN のカテゴリマップを挙げて比較すると、ユニット数が 50 から増えるに従い、同一カテゴリ画像の位相距離が離散的になっていることが確認できる。この結果から、本手法ではコードブックのユニット数を 50 とした。

3.2.2 ART のビジランスパラメータの変化とカテゴリ形成  
本手法において、カテゴリ分類の性能に大きく影響を与えているのが、カテゴリの分類粒度を決める  $\rho$  である。そこで、本手法のカテゴリ分類性能と  $\rho$  の関係を、 $\rho$  を 0.950~0.980 の範囲で 0.005 ずつ変化させて比較した。結果を図 4, 5, 6 に示す。図 4 は  $\rho$  の変化と ART のラベリング結果のグラフ、図 5 は  $\rho=0.950$  と  $\rho=0.975$  のラベリング結果の比較、図 6 は ART のラベリングと CPN のカテゴリマップの関係を示した。なお、コードブックのユニット数を 50、学習画像データ数を 5 枚とした。図 4 より  $\rho = 0.975$ ,  $\rho = 0.980$  のときラベル数が大きく増えているのがわかる。図 5 の ART のラベリング結果を見ると、0.975 でラベル数は 0.950 より多くなったが、カテゴリ 4(face カテゴリ) と他のカテゴリの違いを 0.950 より識別できていることがわかる。また、図 6 の CPN のカテゴリマップを見ると、各画像のマップ配置にはまったく変化がなく、0.975 では 0.950 より多くラベリングされた 7, 8, 9, 10 のユニットが

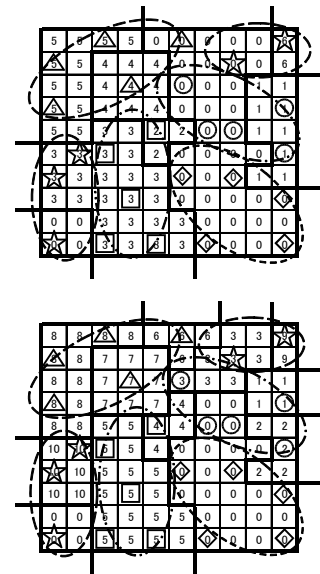


図 6 ART のラベリングと CPN のカテゴリマップの関係 (上: $\rho=0.950$ , 下: $\rho=0.975$ )  
Fig. 6 Relation between labeling by ART and category maps of CPN(Top: $\rho=0.950$ ,Bottom: $\rho=0.975$ ).

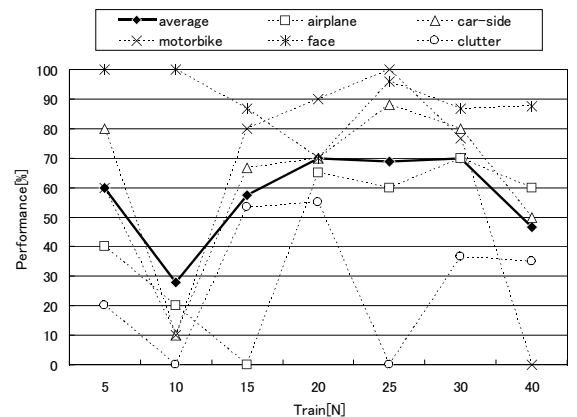


図 7 学習結果における識別率  
Fig. 7 Performance of learning results.

0.950 のマップに重なっただけとなっている。従って、実験結果より、冗長なカテゴリのラベリングを抑制できている ART の性能、CPN の各画像データを各カテゴリで近い位相に写像できる性能の両方が発揮されている  $\rho=0.950$  を本手法で用いる。

### 3.3 評価実験

本節では、学習用画像データ数を 5, 10, 15, 20, 25, 30, 40 枚と変えての評価を行った。テスト用画像データ数は各カテゴリから 25 枚ずつとした [12]。以下に実験結果を示す。

#### 3.3.1 学習画像におけるカテゴリ分類結果

評価実験における学習結果を図 7 に示す。図 7 の平均識別率から学習データ数が 20, 25, 30 枚のときには、識別率が 70% に達している。しかし、一般的に、ニューラルネットワークでは学習画像データ数が増えると識別精度は向上するが、図 7 より、10 枚と 40 枚でそのような結果になっていない。

また、各カテゴリでその分類精度を比較すると、face カテゴリは安定して精度よく識別している。car-side と motorbike カテゴリについても高い分類精度を示しているが、学習画像デー

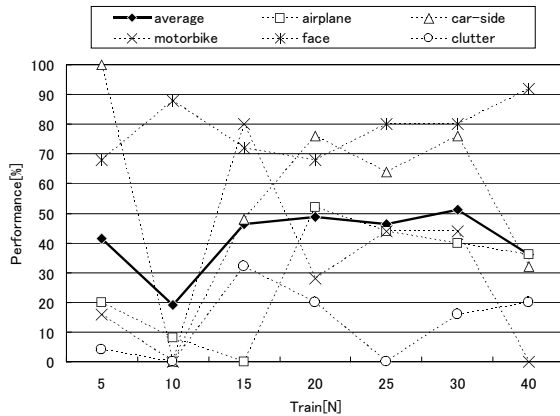


図 8 テスト結果における識別率  
Fig. 8 Performance of test results.

タ数によって大きく精度がばらついている．airplane と clutter カテゴリは学習データ数 10, 15, 25 枚で 0 % を示しており、精度は低いことがわかる．

### 3.3.2 テスト画像におけるカテゴリ分類結果

テスト結果を図 8 に示す．平均識別率より、各学習画像データ数におけるグラフの推移は学習結果の平均識別率のグラフと同じような傾向を示している．最も精度が高かったのは学習画像データ数が 30 枚のときで 51.2 % を示した．

各カテゴリの結果を比較すると、face と car-side カテゴリでは学習結果と同様に学習データ数 5, 20, 25, 30 枚で高い精度を示している．しかし、motorbike カテゴリは学習結果より精度が下がっている．また、clutter カテゴリは学習結果と同様に精度は低いが、airplane カテゴリは学習データ数が多くなるに従い、精度は低いながらもグラフの変化が小さくなって安定してきている．

## 4. 考 察

ART のラベリングと CPN のカテゴリマップについて考察する．airplane の画像は多くの画像で端に白い線が入っている． $\rho = 0.975$  のときは ART のラベリングで、両端で線の太さが異なる画像 2 枚が 0 に、線なしが 1 に、線の太さが両端均等な画像が 2 に、片方だけが 3 とラベリングされており、異なる特徴を捉え、識別できていることがいえる．CPN ではそれらの画像を近い位相に写像しているため、画像の白い線から得られた特徴量を除き、対象物体の特徴を捉えて airplane のカテゴリを形成できたと考えられる．また、motorbike カテゴリでも、3 枚が右向きで 2 枚が左向きとなっており、ART のラベリングで別のカテゴリとしてラベル付けされたが、CPN でカテゴリ形成ができてるのは結果から明らかである．よって、ART と CPN という 2 つのネットワークを unsupervised なカテゴリ分類手法として適用することが有効であるといえる．さらに、入力特徴量としたコードブックに改良を加えることで、その性能が向上すると考えられる．本実験において、ユニット数が 250 のときにカテゴリ形成が離散的になった要因として、ART によるラベル数が少なくなったことが挙げられる．CPN の特性として、入力パターンのラベルは互いに重複しない別々のカテ

ゴリに分かれている必要があり、250 では他に比べて多くの画像でラベルが重複しており、カテゴリ形成にゆがみが生じたために、性能が低下したと考えられる．よって、入力パターンとなるユニットを形成する、SOM のクラスタリングにおける学習回数を増やし、入力パターンの強調を行うことにより ART のパフォーマンスを改善し、カテゴリマップの結果も向上すると考えられる．これは識別する画像枚数が増えたときの細分化にも効果的であると考えられるので、学習画像データ数 40 枚の精度向上にもつながるといえる．

## 5. ま と め

本論文では、ART-2 と CPN を適用した unsupervised なカテゴリ分類手法を提案し、その有効性を確かめた．Caltech256 の 5 種類のカテゴリを用いた実験では、カテゴリ未知な問題に対しても、ART-2 で冗長なカテゴリ変化を抑制し、CPN のカテゴリマップによってカテゴリ形成が可能であることを示した．

今後は、コードブック生成における SOM の学習回数の検討、対象カテゴリ数、データ数が増えたときの、CPN のカテゴリマップの評価方法の検討を行う予定である．さらに、今回の評価実験では本手法の提案と有効性の評価にとどまったが、対象カテゴリ数を増やし、他手法との比較を行う予定である．

## 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] N.Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.
- [3] J.Amores, N.Sebe, and P.Radeva, “Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms”, IEEE, runs. PAMI VOL. 29, NO.10 OCTOBER 2007.
- [4] Y.Freund and R.E.Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting”, Journal of Computer and System Sciences, Aug, 1, Vol.55, pp119-139, 1997.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Hetch-Nielsen, R. 1987. Counterpropagation networks, Proc. of IEEE First Int'l. Conference on Neural Networks. 1987. :19-32.
- [9] 藤吉弘巨, “Gradient ベースの特徴抽出 - SIFT と HOG - ”.
- [10] T.Kohonen, Self-Organizing Maps, Springer Series in Information Sciences, 1995.
- [11] Carpenter, G.A. and Grossberg, S., “Pattern Recognition by Self-Organizing Neural Networks”, The MIT Press, 1991.
- [12] Greg Griffin, Alex Holub and Pietro Perona, “Caltech-256 Object Category Dataset”, March 14, 2007.