

ロボットビジョンのシーン記述における SIFT 特徴点の選択

内海 祐哉[†] 間所 洋和[†] 佐藤 和人[†]

[†] 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: †{B08A018,madokoro,ksato}@akita-pu.ac.jp

あらまし 本論文では, OC-SVM (One Class-Support Vector Machines) を用いて, 一般物体認識に有効な SIFT (Scale-Invariant Feature Transform) 特徴点を選択する手法を提案する. 我々は, ART-2 (Adaptive Resonance Theory-2) と CPN (Counter Propagation Networks) を組み合わせた教師なしカテゴリ分類法を提案した. 従来手法では非注目領域の特徴点が分類に影響を与える. 提案手法では検出された特徴点から注目領域の特徴点を選択することにより, カテゴリ分類に有効な特徴記述をする. 評価実験では, Caltech-256 データセットと移動ロボットから得られる時系列画像のカテゴリ分類において, 従来手法と比べて分類精度が向上し, ロボットビジョンにおけるカテゴリ分類に有効であることを示す.

キーワード SIFT, OC-SVM, ART-2, CPN

Selection of SIFT Feature Points for Scene Description in Robot Vision

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

[†] Faculty of Systems Science and Technology, Akita Prefectural University 84-4 Aza Ebinokuchi Tsuchiya,
Yurihonjo City, 015-0055 Japan
E-mail: †{B08A018,madokoro,ksato}@akita-pu.ac.jp

Abstract This paper presents a method for selection of SIFT(Scale-Invariant Feature Transform) feature points using OC-SVM (One Class-Support Vector Machines). We proposed the method for automatic category classification using a network system that combine incremented learning function of ART-2(Adaptive Resonance Theory-2) networks and self-presentation characteristic of CPN (Counter Propagation Networks). In our former method, the feature points of nontarget region make misclassification. In our method, OC-SVM selects feature points of target region. Experiment results that used Caltech-256 object category dataset and the time-series image dataset caught from a mobile robot show that classification accuracy of our method is better than that of the former method.

Key words SIFT, OC-SVM, ART-2, CPN

1. はじめに

ロボットが人間の生活環境である実環境において様々なタスクを達成するためには, 環境を理解し自律的に行動することが必要不可欠である. ロボットの自律行動に必要な要素技術の一つとして, 時々刻々と変化する実環境において能動的に周辺の状態を認識しながら概念パターンとなる世界像を獲得することが求められる. 世界像を獲得するには, 視覚情報から場所や場面状況などのシーンにおける注目領域を記憶し, それをカテゴリごとに分類する必要がある.

計算機の発展により大量のデータを高速に処理可能となり, 機械学習を物体認識に適用できるようになってきた[1]. 物体認識において機械学習を用いる手法では, 学習用の画像中に教師情報が含まれる教師ありカテゴリ分類と, 教師情報が含ま

れない教師なしカテゴリ分類がある. 実環境をロボットが自律的に行動し, 環境を理解するためには, 教師なしカテゴリ分類が重要となる. 教師なしカテゴリ分類の代表的な手法として, 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, 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化を行っている[2]. 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[3]. これらの手法では, いずれもカテゴリ数が既知でなければならない. このようなカテゴリ数が既知のカテゴリ分類は, カテゴリ数を一意に決められない実環境には対応が困難のため, ロボットビジョンの分野では, カテゴリ数が

未知の教師なしカテゴリ分類が必要となる。

画像上の分類に直結する領域を注目領域、それ以外の領域を非注目領域とすると、教師なしカテゴリ分類は背景領域（非注目領域）の影響を受けやすいため、オブジェクト領域（注目領域）の特徴量選択が重要である。カテゴリ分類の特徴抽出方法として、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を記述する SIFT (Scale-Invariant Feature Transform) がよく用いられている [4]。SIFT の処理手順は、特徴点の検出と特徴点の記述の 2 段階からなる。特徴点の検出では、DOG (Difference-of-Gaussian) 処理によりスケールとキーポイントを検出し、検出されたキーポイントから主曲率とコントラストにより有効な特徴点に絞り込み、サブピクセルの推定を行う。特徴点の記述では、回転に不変な特徴を得るためにオリエンテーションを算出し、求めたオリエンテーションに基づいてキーポイントの特徴量を記述する。通常は画像全体から特徴量を抽出しているため、注目領域と非注目領域という概念がなく、SIFT を用いる手法では、注目領域だけでなく非注目領域を含んだ特徴量を記述する。カテゴリ分類の精度を向上させるためには、注目領域と非注目領域に分けて、それらの関係性も利用し、記述することが望ましい。注目領域の特徴量による記述を実現する手法として、物体と背景の境界情報を用いて、物体領域のみの情報による特徴点の検出と特徴量を記述する SIFT を拡張した BSIFT (Background and Scale-Invariant Feature Transform) がある [5]。しかし、BSIFT では、境界情報が必要であり、物体の境界情報をあらかじめ求めたシーンを対象とするため、物体領域と背景の境界情報を一意に決められない実環境では適用が困難である。

人間は情報が集中している箇所に注目している。例えば、複雑な背景中にイスがある場合には、自然とイスに注目する。このように画像上で注目する領域は、シーンを特徴付ける重要な領域の一つと考えられる。そこで、本研究では、実環境におけるシーンでの有効な特徴点選択を、注目領域か非注目領域の 1 クラス識別問題と考え、注目領域の特徴点（注目点）を選択するため、OC-SVM (One Class-Support Vector Machines) を用いる。OC-SVM は密度関数の推定を行わずに、未知のパターンでの高密度領域の推定を行える教師なし学習である。この OC-SVM によって選択された特徴量を、我々が提案している教師なしカテゴリ分類法に適用し、カテゴリ分類の実験をする。評価実験ではカテゴリ分類研究において一般的に用いられる Caltech-256 データベースによるカテゴリ分類と、ロボットビジョンの視覚から得られるから視野画像列を用いたカテゴリ分類について報告する。

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。3. では Caltech-256 データセットと移動ロボットを用いる実験方法と結果を示し、4. でその考察をする。最後に 5. でまとめと今後の課題について述べる。

2. 提案手法

本節では、提案するカテゴリ分類手法について述べる。我々は、時間軸に沿った学習を進めることができる ART-2(Adaptive

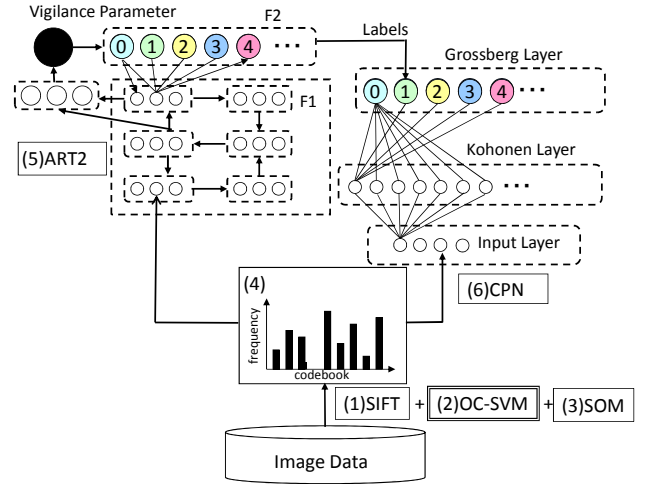


図 1 ネットワーク構成

Fig. 1 Network architecture of the proposal method.

Resonance Theory-2) [6] と CPN(Counter Propagation Networks) [7] の自己写像特性を組み合わせたカテゴリ数が未知な問題にも適用できる教師なしカテゴリ分類法（以下、従来手法と記す）を提案した [8]。従来手法では画像全体で SIFT 特徴量を記述しているため、非注目点がカテゴリ分類に負の影響を与えていると考えられる。したがって、注目点と非注目点に分けて特徴量を記述する必要があり、本手法では、カテゴリ分類に有効な特徴量記述をするために、OC-SVM で注目点を選択し、カテゴリ分類を試みる。

2.1 提案手法の処理手順

図 1 に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示す。我々の手法の処理手順は、(1)SIFT による特徴点検出および SIFT 記述子ベクトルの算出、(2)OC-SVM による特徴点選択、(3) 全学習画像の SIFT 記述子ベクトルの SOM 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の SIFT 記述子ベクトルのヒストグラムを作成、(5)ART-2 による教師なしラベルの形成、(6)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。従来手法では、画像全体で非注目点を含む SIFT 特徴量を記述しているため、非注目点がカテゴリ分類に影響を与えている。本手法は教師なしカテゴリ分類の精度を向上させるため、コードブック作成の前に、OC-SVM を組み込み、カテゴリ分類に有効な特徴点を選択する。

2.2 OC-SVM による特徴点選択

SVM [9] を拡張した OC-SVM 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である。RBF(Radial Basis Function) カーネルで特徴空間へ写像をすると、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。この性質を利用し、原点近くと他のデータを識別する超平面を設定する。超平面の位置は、パラメータ ν によって変わり、パラメータ 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。

識別関数を

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

としたとき, 制約条件

$$\begin{aligned} \omega^\top z_i &\geq \rho - \zeta_i, \quad i = 1, \dots, l \\ \zeta_i &\geq 0, \quad i = 1, \dots, l, \quad 0 < \nu \leq 1 \end{aligned} \quad (2)$$

式 (2) のもとで, 次の最適化問題を解く.

$$\frac{1}{2} \|\omega\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \zeta_i - \rho \rightarrow \omega, \zeta, \rho \text{ を最小化} \quad (3)$$

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.

$$\Phi: x_i \mapsto z_i \quad (4)$$

式 (3) の最適化問題の解を ω と ρ とする. パラメータ ν は, 学習データ z_i に対し, 式 (1) の識別関数が正の値をとり, $\|\omega\|$ は小さな値をとるように制御する. 式 (3) のラグランジュ関数を計算すると,

$$\begin{aligned} L(\omega, \zeta, \rho, \alpha, \beta) &= \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \zeta_i - \rho \\ &\quad - \sum_{i=1}^l \alpha_i (\omega^\top z_i - \rho + \zeta_i) - \sum_{i=1}^l \beta_i \zeta_i \end{aligned} \quad (5)$$

となり, 最適化問題を解くために, 双対最適化問題を導出する. 式 (3) のラグランジュ関数の α_i と β_i について最大化し, ω, ρ, ζ については最小化する. 最適解において, クーン・タッカー条件より, パラメータ ω, b, ρ, ζ でラグランジュ関数 L を偏微分すると, 0 になるので,

$$\frac{\partial}{\partial \omega} L = 0 \rightarrow \omega = \sum_{i=1}^l \alpha_i z_i \quad (6)$$

$$\frac{\partial}{\partial \zeta_i} L = 0 \rightarrow \alpha_i = \frac{1}{\nu l} - \beta_i \quad (7)$$

$$\frac{\partial}{\partial \rho} L = 0 \rightarrow \sum_{i=1}^l \alpha_i = 1 \quad (8)$$

クーン・タッカー条件により, 鞍点で,

$$\left. \begin{aligned} \alpha_i \cdot [\rho - \zeta_i - \omega^\top z_i] &= 0, \quad i = 1, \dots, l \\ \rho - \zeta_i - \omega^\top z_i &\leq 0, \quad i = 1, \dots, l \\ 0 \leq \alpha_i &\leq \frac{1}{\nu l}, \quad i = 1, \dots, l \\ \beta_i \cdot \zeta_i &= 0, \quad -\zeta_i \leq 0, \quad \beta_i \geq 0, \quad i = 1, \dots, l \end{aligned} \right\} \quad (9)$$

式 (5) のラグランジュ関数に式 (6) ~ 式 (9) を代入し, 内積をカーネルに置き換えると,

$$\frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j k(z_i^\top z_j)$$

$$0 \leq \alpha_i \leq \frac{1}{\nu l}, \quad i = 1, \dots, l, \quad \sum_{i=1}^l \alpha_i = 1 \quad (10)$$

双対最適化問題を得ることができる. 式 (9) の条件を満たし, かつ $\alpha_i > 0$ と $\zeta_i = 0$ の条件を満たす学習データ z_i をサポートベクタとする.

式 (6) の展開式から識別関数は,

$$f(z) = \sum_{i=1}^l \alpha_i k(x_i, z) - \rho \quad (11)$$

となる. ただし, $\zeta_i = 0$ である.

最適解で α_i, β_i が 0 でないなら, 以下の等式が成立する.

$$\sum_{j=1}^l \alpha_j k(x_j x_i) = \rho \quad (12)$$

これにより, ρ を計算する.

カーネルトリックを利用することで, 双対問題となる識別関数に, 高次元特徴空間上で点 $\Phi(x)$ が現れないため, 高次元特徴空間での膨大な内積計算を避けることができ, 計算量を削減できる.

式 (3) の最適化問題の解が, $\rho \neq 0$ を満たす場合, OC-SVM 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. 実験結果

本節では評価画像のデファクトスタンダードとなっている Caltech データセットと移動ロボットから得られる視野画像列

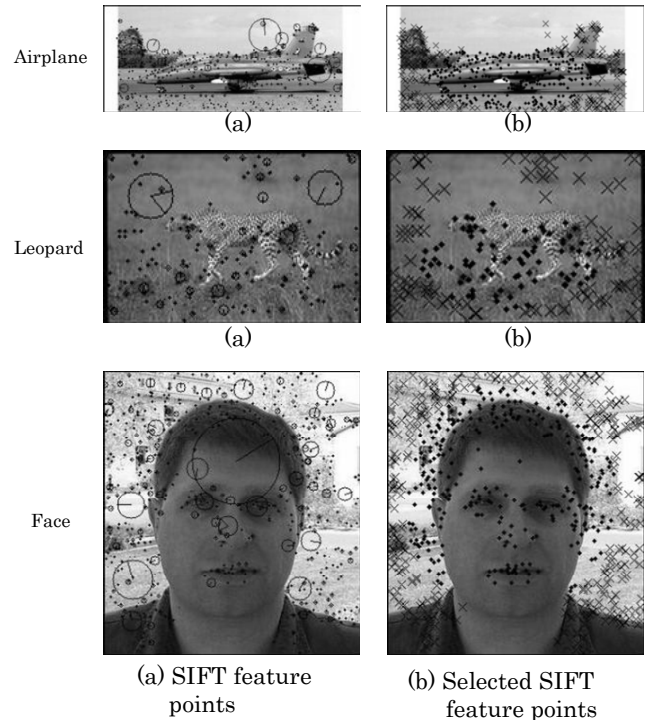


図 2 特徴点選択結果 (●:注目点, ×:非注目点)

Fig. 2 Selected feature points

(●:Selected points, ×:Other points).

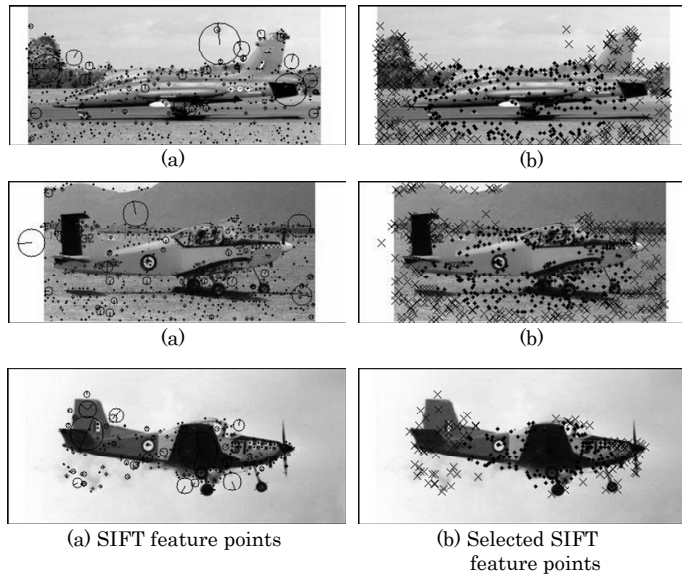


図3 特徴点選択結果 (●:注目点 ×:非注目点)

Fig. 3 Selected feature points

(●:Selected points, ×:Other points).

表1 従来手法と提案手法カテゴリ分類率

Table 1 Rate of classification.

	Data	5category	10category	20category
Our former method	Learning	84 %	70 %	64 %
	Test	76 %	30 %	38 %
Our method	Learning	96 %	94 %	81 %
	Test	76 %	44 %	50 %

を用いた実験を行った。Caltech-256 データセットと視野画像列のカテゴリ分類において、OC-SVM による特徴点の選択を行うことにより、本手法のカテゴリ分類での有効性を示す。

3.1 Caltech-256 の画像を用いた実験

本実験では、Caltech-256 データセット (256 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。5 カテゴリ、10 カテゴリ、20 カテゴリでカテゴリ分類実験から、本手法における注目点を評価した。図 2 に異なるカテゴリの特徴点選択結果、図 3 に Airplane カテゴリの特徴点選択結果、図 4 に 5 カテゴリにおける ART-2 のラベル付け結果 (縦軸: ラベル番号, 横軸: データ数)、図 5 に 5 カテゴリにおける CPN のカテゴリマップ、表 1 に従来手法と提案手法の分類率を示す。なお OC-SVM のパラメータ を 0.50, SOM のユニット数を 100, 学習回数を 100,000 回, ART-2 のビジランスパラメータを 0.930, CPN の学習率係数 $\alpha(t)$ と $\beta(t)$ を 0.5, 学習回数を 10,000 回とした。

図 2 に示す特徴点選択結果から、カテゴリごとに異なる注目点を選択していることがわかる。図 3 に示す Airplane カテゴリでは、後方の翼を含むオブジェクト全体領域ではなく、Airplane の中でも中央付近を注目点として選択している。図 4 の ART-2 の分類結果から、従来手法は Leopard カテゴリはラベルが混同しているが、本手法は混同せずにラベル付けをし、ラベルの混同が減少した。図 5 の CPN のカテゴリマップの Airplane に注目すると、有効な特徴点を選択することにより、Airplane の画

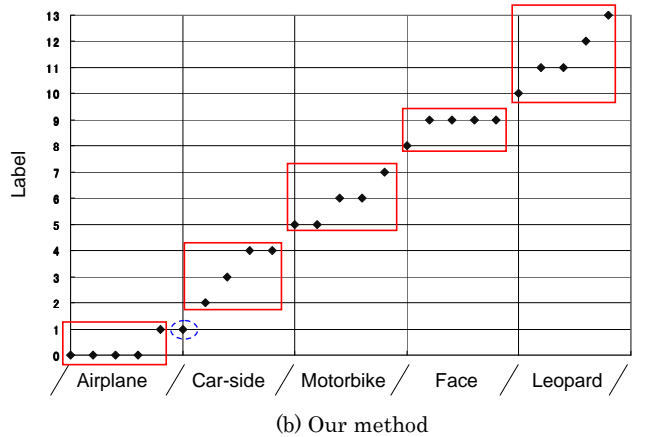
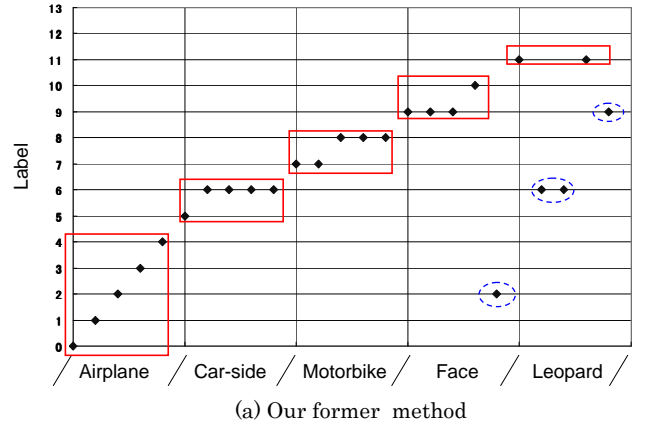


図4 5 カテゴリにおける ART-2 のラベル付け結果

Fig. 4 Labeling results of ART-2 at 5 categories.

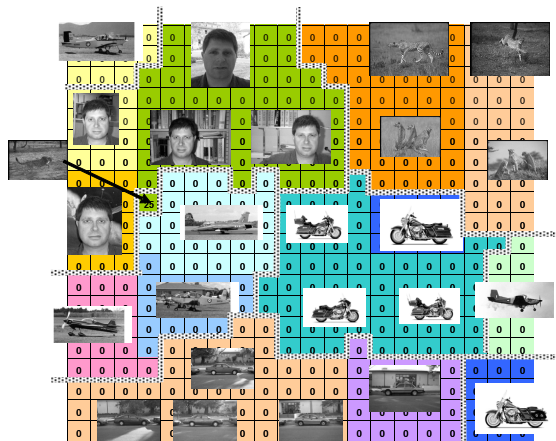
像を隣接するユニットに写像することができた。また、ART-2 でラベルが混同した Car-side カテゴリの画像は同じカテゴリへ写像されている。表 1 から、OC-SVM による特徴点の選択を行うことにより、いずれのカテゴリ数においても分類率が向上したことから、OC-SVM によりカテゴリ分類に有効な注目点の特徴点を選択できている。

3.2 移動ロボットを用いた実験

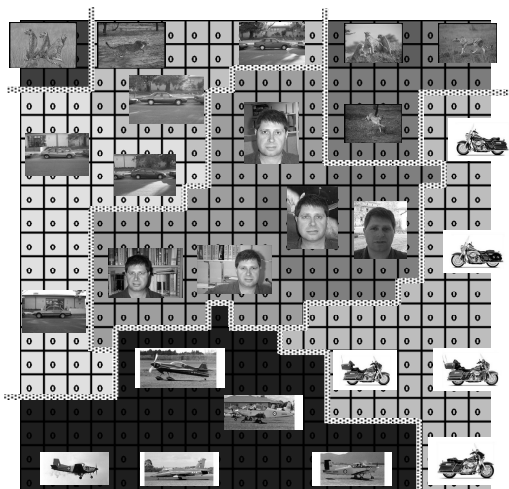
本実験では、図 6 に示す本学の廊下を走行したときの全方位カメラから得られる視野画像列を用いてカテゴリ分類実験を行った。実験で扱うカメラは、1/3 "interline CCD, 解像度が 640 × 480pixel, フレームレートが 30fps の全方位カメラである。ロボットのカメラの高さは 1,500mm, 平均移動速度が約 30m/min, 廊下の幅は 1,830mm である。

図 7 に全方位カメラから得られる時系列画像の特徴点選択結果、図 8 に ART-2 によるラベル付け結果、図 9 に CPN のカテゴリマップを示す。図 6 において、をスタートし、反時計回りに一周する。走行経路において、シーンの見え方が類似する区間を A~H としてカテゴリ分類の目安とする。なお、OC-SVM のパラメータ を 0.82, ART-2 のビジランスパラメータを 0.850 とした。

図 8 の ART-2 のラベル付け結果から、特徴を選択することによりラベルの混同が少なくなりラベル数も減少した。図 9 の CPN のカテゴリマップから、特徴選択しない場合は 8 カテゴリ



(a) Our former method



(b) Our method

図5 5カテゴリにおけるCPNのカテゴリマップ
Fig.5 Category map of CPN at 5 categories.

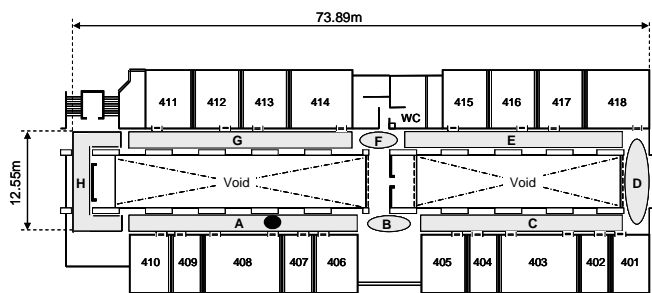
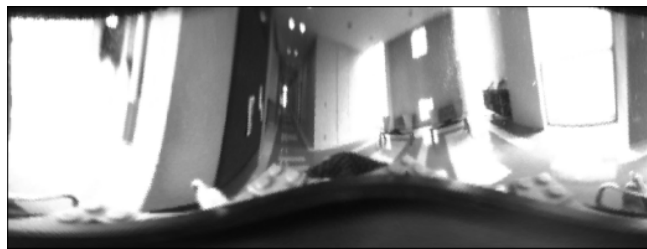


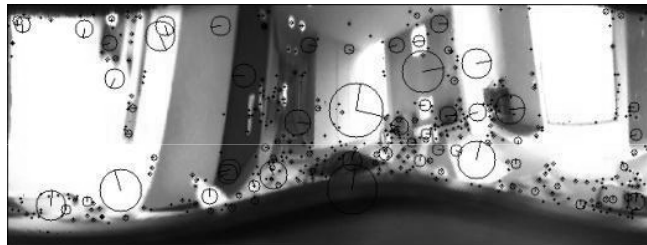
図6 実験環境

Fig.6 Experimental Enviroment.

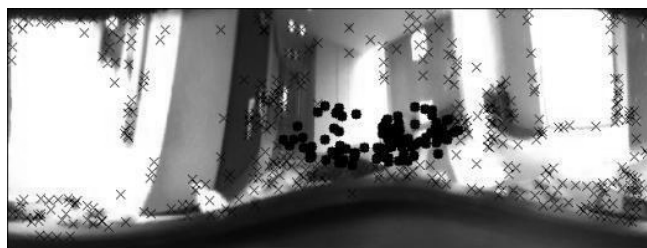
中4カテゴリしか形成できず、同じカテゴリを隣接するユニットに写像できていないが、特徴を選択することにより、8カテゴリ中すべてのカテゴリを形成でき、類似したシーンを隣接するユニットに写像できている。シーンごとのART-2のラベル付けをもとに、CPNで類似したシーンを隣接するユニットに写像し、走行範囲でコーナーと直線のカテゴリを形成できている。これらのことから、OC-SVMで選択された注目点はロボットビジョンにおけるカテゴリ分類に有効であるといえる。



(a) Original Image



(b) SIFT feature points



(c) Selected SIFT feature points

図7 視野画像の特徴点選択結果(●:注目点, ×:非注目点)

Fig.7 Selected feature points

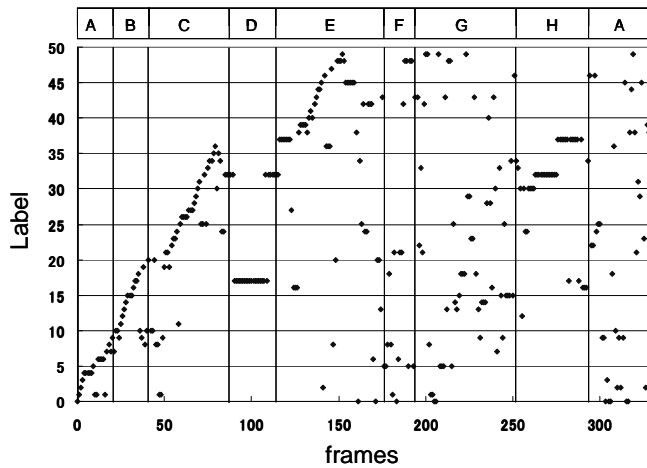
(●:Selected points, ×:Other points).

4. 考 察

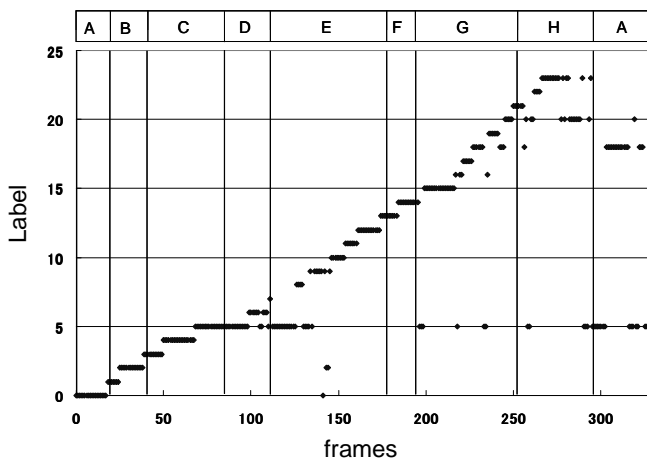
Caltech-256を用いた実験において、人間がオブジェクトを分類するとき、オブジェクト全体から判断しているわけではなく、オブジェクトを特徴付ける情報が集中している箇所に注目している。本手法のOC-SVMは、オブジェクト全体を注目点として選択しているのではなく、オブジェクトの中でも情報が集中している箇所を注目点として選択している。これにより、カテゴリ分類に有効な特徴記述ができ、分類精度が向上したと考えられる。

ロボットの視覚から得られる視野画像列を用いた実験では、ART-2のラベル付け結果より、本手法はラベル5とラベル20においてラベルの混同が多かった。ラベル5とラベル20はコーナー付近で、窓が多く明るい画像である。混同している画像は、窓と部屋のドアが開いていることによる光で明るく、窓やドアの枠の特徴を選択している画像である。つまり、ラベル5とラベル20で選択された特徴が類似していることにより、ラベルが混同したと考えられる。

本手法のCPNのカテゴリマップから、廊下の直線部分であるC, E, Gカテゴリが隣合っている。これは、同じようなカテゴリを近くに写像できているといえる。また、図面上でHカテゴリとDカテゴリはコーナー付近で同じようなカテゴリであるが、Hカテゴリには大きな窓があり、マップ上で遠くに写像



(a) Our former method



(b) Our method

図 8 ART-2 のラベル付け結果

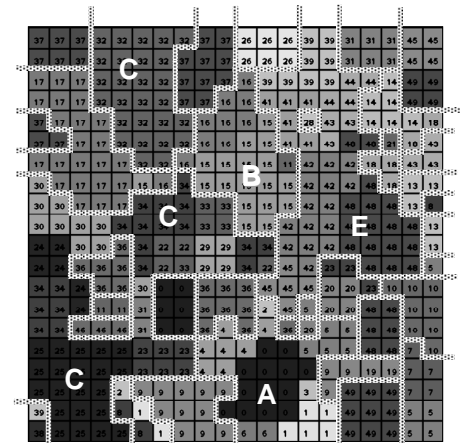
Fig. 8 Labeling results of ART-2.

できている．以上のことから，ART-2 のラベルをもとに CPN によってカテゴリ間の空間的関係性を写像できている．

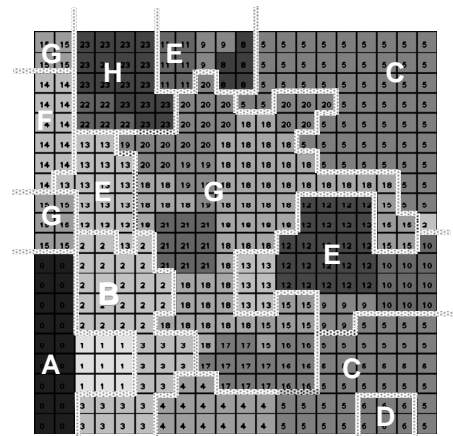
5. まとめ

本研究では，OC-SVM を用いて，SIFT 特徴点を選択する手法を提案し，Caltech-256 のデータセットとロボットの視覚から得られる時系列画像を用いた実験から本手法の有効性を示した．Caltehe-256 の画像を用いた実験では，学習データにおけるカテゴリ分類率が 80 % 以上であることから，OC-SVM で選択した注目点が分類精度の向上に効果がある．ロボットの視覚から得られる視野画像列において，従来手法ではシーンごとにラベルを付けることができず混同したが，本手法は時間軸に沿ってラベル付けを行うことができ，ロボットビジョンにおけるカテゴリ分類に有効であることがわかった．

今後の課題として，注目点と非注目点の特徴量の関係性をコンテキストとしてカテゴリ分類へ利用することと，オリエンテーションなどの他の SIFT 特徴量も有効に利用することが挙げられる．また，時系列画像を ART-2 が時間軸に沿った学習しているように，OC-SVM で時間軸の連続性を活用を検討する予定である．



(a) Our former method



(b) Our method

図 9 CPN のカテゴリマップ

Fig. 9 Category map of CPN.

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] Long(Leo) Zhu, Yuanhao Chen, and Alan Yuille, "Unsupervised Learning of Probabilistic Grammar-Markov Models for Object Categories," IEEE, Truns. PAMI VOL.31, NO.1, Jan. 2009.
- [3] 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.
- [4] 藤吉弘亘, "Gradient ベースの特徴抽出 - SIFT と HOG -," CVIM, [コンピュータビジョンとイメージメディア] 2007, pp.211-224, 2007.
- [5] A.Stein, M. Herbert, "Incorporating background invariance into feature-based object recognition," Proc. of IEEE Workshop on Applications of Computer Vision and Pattern Recognition (CVPR), pp.1978-1983, 2006.
- [6] Carpenter, G.A. and Grossberg, S., "ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Ppterns," Applied Optics, Vol.26, pp.4919-4930, 1987.
- [7] Hetch-Nielsen, R., "Counterpropagation networks," Proc. of IEEE First Int'l. Conference on Neural Networks. 1987.
- [8] 塚田真敬, 間所洋和, 佐藤和人, "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," PRMU2008-148, pp.7-12, Dec. 2008
- [9] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning Vol.20, No.3, pp.273-297 1995.