

H-046

## 一般物体認識におけるクラス内変化を考慮した 認識手法に関する基礎研究

An experimental study on a recognition method for considering intra-class  
variability in generic object recognition

下地竜雄馬<sup>†</sup> 當間愛晃<sup>‡</sup> 赤嶺有平<sup>‡</sup> 山田孝治<sup>‡</sup> 遠藤聡志<sup>‡</sup>  
Ryouma Shimoji Naruaki Toma Yuhei Akamine Koji Yamada Satoshi Endo

### 1. はじめに

高速データ通信や大容量記憶装置の普及に伴い、画像・映像データ量は爆発的に増加している。そのため、画像・映像情報そのものの意味に即した認識・検索技術が必要とされている。しかし、画像・映像から得られる情報では意味そのものに直接結びつかないため、意味に即した認識・検索は非常に困難だとされている。この問題をセマンティックギャップと呼び、この問題を克服するための試みとして一般物体認識の研究が盛んに行われている [1]。

一般物体認識とは、制約のない実世界シーン画像に対して計算機がその中に含まれる物体を一般的な名称で認識することである。一般物体認識における課題として、特定の物体検出 (Localization) 問題と画像分類 (Categorization) 問題の2つが重要な問題とされている。Localization 問題とは、対象とするクラスが画像中のどこにあるのかを認識することであり、Categorization 問題とは、その画像がどのクラスに属するかを認識することである。これらの問題が一般物体認識における重要な課題とされ、一般物体認識を行うためにはこれらの問題を解決する必要がある。

近年では、一般物体認識の手法として、SIFT (Scale-Invariant Feature Transform) 特徴量 [2] を用いた BoF (Bag-of-Features) [3] による認識手法が注目を浴びている。この手法は、一般物体認識手法において高い精度を得ているが、認識率は対象クラス毎に大きく異なり、安定した認識精度を実現できていない。その原因の1つとしてクラス内の変化が挙げられる。例えば「椅子」の場合、4本足の椅子であったり、1本足の椅子、またはソファやベンチなどと様々な種類が挙げられる。人間ならば全てを「椅子」だと認識できるが、コンピュータビジョンの世界ではこれを「椅子」だと認識することは非常に困難とされている。そのため、本研究ではクラス内変化を考慮に入れた一般物体認識の手法として、1クラスにおける学習用画像において、クラスタリングを行うことでクラス内変化を捉えた認識

を行う。

### 2. 特徴量抽出手法

#### 2.1. BoF (Bag-of-Features)

BoF とは、文書分類手法である Bag-of-Words を画像に適応させたものであり、画像を局所特徴の集合とみなすことで局所特徴の位置を考慮せず出現頻度により画像を表現し、その認識を行う。

BoF における処理は、学習と認識に分類される。画像を局所特徴量の集合とみなす BoF の学習では、まず全学習用画像から回転・スケール・照明変化にロバストな SIFT 特徴量の抽出を行う。その後、得られた全ての SIFT 特徴量を特徴空間において k-means クラスタリングを行うことにより visual words の生成を行う。各画像から得られた SIFT 特徴量を、生成した visual words から最近傍の visual word に割り当てることでベクトル量子化を行い、visual word 次元のヒストグラムにより表現される。画像毎に抽出される特徴量は異なるため、特徴量の数により正規化し、SVM による学習を行う。

認識では、学習と同様に画像をベクトル量子化ヒストグラムで表現し、SVM の学習モデルでの認識を行う。

#### 2.2. HOG (Histogram of Oriented Gradients) 特徴量

HOG 特徴量 [4] とは、SIFT 特徴量と同様に局所領域における輝度の勾配方向をヒストグラム化した特徴量である。SIFT 特徴量と類似した特徴量の記述を行うが、特徴点に対して特徴量を記述する SIFT 特徴量に対し、HOG 特徴量ではある一定領域に対する特徴量の記述を行う。そのため、大まかな物体形状を表現することが可能であり、得られた特徴量も用いた統計的学習手法による物体検出などに使用される。

HOG 特徴量は、画像をセル単位で分割し、各セルに対してその領域内における勾配方向ヒストグラムを生成する。勾配方向は  $0^\circ$  から  $180^\circ$  を範囲とした9方向に量子化される。その後、各セルにおいて求められた勾配方向ヒストグラムを用いて  $3 \times 3$  セルを1ブロックとした領域内での正規化を行う。その際、ブロックを

<sup>†</sup> 琉球大学大学院理工学研究科情報工学専攻

<sup>‡</sup> 琉球大学工学部情報工学科

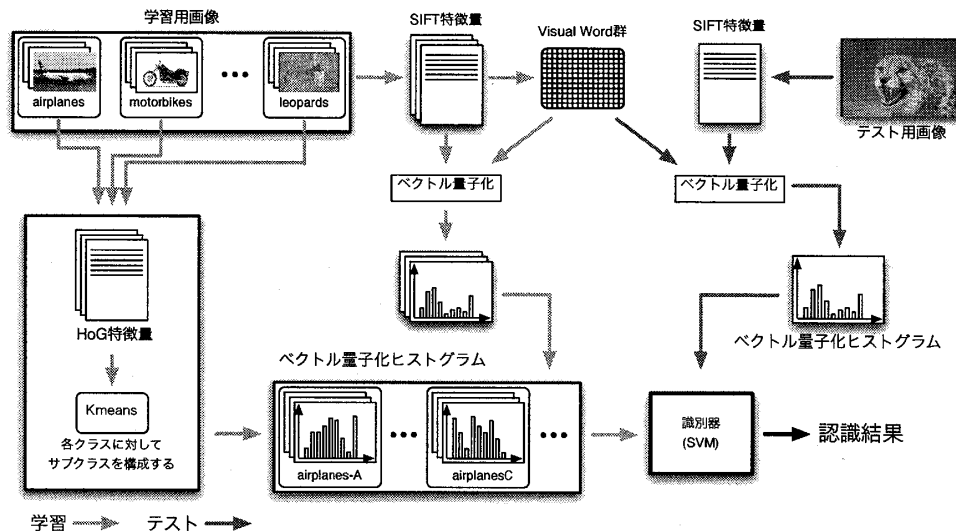


図 1: 全体の構成

1セルずつ移動しながら正規化を行うため、局所的な照明変化にロバストな特徴量を得ることができ、画像全体の形状特徴を表現することが可能である。

### 3. 提案手法

本研究では、一般物体認識におけるクラス内変化を考慮した認識を行う。クラス内変化を考慮するため、各クラスに対してサブクラスの構成を行った上での学習を行う。各クラスに対してサブクラス化を行う際には、クラス内変化をどう捉えるかが重要となる。本研究では、形状特徴を用いることで見方・形状・種類の変化に対応したサブクラス構成できると考え、HOG 特徴量を用いたサブクラス化を行い、それぞれに識別器を構成することでクラス内変化に即した認識を行う。図 1 に全体の構成を示す。

#### 3.1. 学習

Step1: HOG 特徴量を用いたサブクラス化 (ラベル生成)

全学習用画像から HOG 特徴量を抽出する。その後、各クラスに対して k-means クラスタリングによりサブクラス化を行う。その際に HOG 特徴量の次元数の大きさがネックとなるため、HOG 特徴量のバイナリ化 [5] を行い、HOG 特徴量の削減を行った上でクラスタリングを行う。ここで得られたクラスを Step3 におけるラベルとして利用する。

Step2: SIFT 特徴量を用いた BoF によるヒストグラム化 (特徴ベクトル生成)

全学習用画像から SIFT 特徴量を抽出し、visual words を生成する。生成された visual word により各画像をベクトル量子化ヒストグラムとして表現し、それを Step3 における特徴ベクトルとして利用する。

Step3: SVM による識別器の構成

SVM による学習では、Step1 により得られたサブクラス化情報を用いてラベル付けを行い、サブクラス毎に識別器の構築を行う。得られた学習データにより、マルチクラス SVM である LIBSVM<sup>§</sup> を用いて学習を行う。SVM による学習ではカーネルの選択も重要となり、線形・多項式・RBF・シグモイドの 4 種類の Kernel で実験し、最も精度の高かった線形 (Linear) を用いることとする。

#### 3.2. 認識

認識では、学習過程の Step2 により生成された visual words により画像をベクトル量子化ヒストグラムで表現する。その後、構築された識別器による認識を行う。また、学習過程においてサブクラス毎にラベル付けを行ったので、認識後にどのクラスかを判断して認識結果を出力する。

### 4. 実験

#### 4.1. 実験方法

実験では、Caltech101 を用いたデータセットの 5 分割交差検定法により評価を行うことで、学習・テストデータセットによる偏りを防ぐことができる。

<sup>§</sup>LIBSVM : <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

サブクラス化を行うにあたり、クラスタリングの手法として k-means クラスタリングを用いる。k-means クラスタリングでは初期クラスの状態により最終結果が大きく依存される。そのため、各クラスに対して10回のクラスタリングを行い、それぞれの認識率の平均を求めると結果の揺らぎを防ぐ。また、k-means クラスタリングではクラス数が固定されるため、クラス数を2~7の間で実験を行ない、各クラスにおいて最も精度の良かったクラス数を認識結果とする。

実験結果の評価基準には適合率・再現率を用いた総合的な評価として F 値を用いる。

最終認識結果は、上記で述べたような認識率の平均と、各クラスにおける最高認識率により評価を行う。また、実験の際に複数回行ったクラスタリングの精度が認識結果にどのような影響を与えるかの検証を行い、HOG 特徴量がクラス内変化を捉えるための情報として適切かどうかの検証を行う。

#### 4.2. データセット

実験には、一般物体認識におけるアルゴリズム製法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、カリフォルニア工科大学の Caltech101<sup>¶</sup> を用いた。

Caltech101 のデータセットから 10 クラスをランダムに選択 (表 1) し、各カテゴリーに対して 50 枚の画像を用意した。

表 1: Caltech101

Class 番号	Class 名	Class 番号	Class 名
1	airplanes	6	leopards
2	chair	7	llama
3	cup	8	motorbikes
4	grand_piano	9	schooner
5	joshua_tree	10	sunflower

#### 4.3. 実験結果

まず、BoF による認識率と本研究における提案手法の平均認識率・最高認識率の結果を以下 (表 2, 図 2) に示す。

表 2: 全クラスを平均した実験結果 (%)

BoF による認識手法	65.92
提案手法 (Average)	65.95
提案手法 (Max)	77.35

表 2 は、全クラスにおける認識率を平均化した結果

を示している。この結果を見ると、提案手法の平均では既存の BoF による認識手法と大差ないが、提案手法における最高認識率では既存の手法を大きく上回っている事がわかる。しかし、この結果からは各クラスがどのような認識率になっているかが判断できないため、図 2 にそれぞれの認識率のグラフを示す。

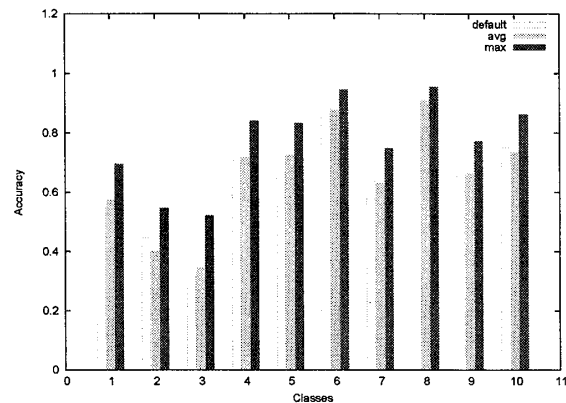


図 2: 実験結果

この結果から、提案手法における平均認識率では誤差だと言えるが、最高認識率においては全クラスにおいて精度が上がっていることが分かる。最高認識率は、実験方法でも述べたように複数回クラスタリングを行った上で、そのクラスにおいて最も精度の高かった認識結果である。このことから、サブクラス化を行うにあたりクラスタリングの精度を上げることが出来れば、平均認識率においても安定した認識率を出すことができると考えられる。

k-means クラスタリングでは、実験方法でも述べたように初期状態により最終結果が大きく変化するため、評価関数を設けることで複数回の実行から優秀な結果を使用する対処法が一般的に用いられる。その評価関数として式 (1) が設けられており、Score が小さいほど良いクラスタリング結果だといえる。 $D(x, c_i)$  は要素  $x$  と  $x$  が属するクラスターの重心  $c_i$  との距離である。

$$Score = \sum_{i=c}^k \sum_{x \in C_i} (D(x, c_i))^2 \quad (1)$$

( $c_i$ : クラスターの重心)

上記の式 (1) により、実験結果で得られたそれぞれのクラスタリング結果を評価し、その Score と認識率において相関が見られるかを検証した。その結果の一部を以下の図 3 に示す。

上記の実験結果は Dataset1 の motorbikes クラスにおけるクラス数毎の分布を plot した結果である。こ

<sup>¶</sup>Caltech101  
[http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)

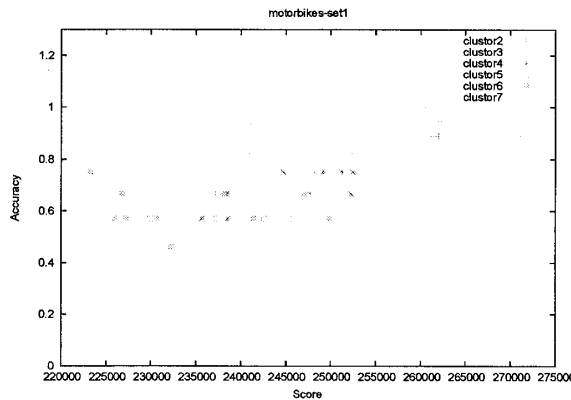


図 3: motorbikes の相関図

こでの Dataset1 とは、5 分割交差検定法により分割した 1 パターンである。学習用画像によって適切なクラスタ数は変化するため、パターンごとに plot して相関を見る必要がある。横軸が評価関数における Score、縦軸がそのクラスタリング結果を用いた認識率となっており、色の違いはそれぞれクラスタ数となっている。

この結果を見て分かる通り、クラスタリング精度の指標となる Score と認識率には際立った相関関係は見られなかった。また、クラスタリングには背景の有無によりクラスタリングの精度は大きく異なり、背景のない motorbikes クラスでは他クラスと比較すると、上手くクラスタリングされているように見えた。HOG 特徴量は、画像全体から特徴量の抽出を行うので、ノイズとなる情報も大量に含まれている。そのため、今回はバイナリ化による HOG 特徴量の削減を行った上でのクラスタリングを行ったが、それでも安定した精度を出すことが出来なかった。

次に、各クラスに対して行ったクラスタリングを BoF により得られたベクトル量子化ストグラムを用いて行なってみた認識結果を図 4 に示し、HOG 特徴量を用いてサブクラス化を行った認識結果との比較を表?? に示す。

この結果から、平均・最高認識率共に HOG 特徴量を用いてサブクラス化を行った認識結果 (図 2) と比較し、一定の効果しか得られなかった事が分かる。また、各クラスタリング結果の Score 分布を確認すると、ほぼ同じようなクラスタリングとなっていることが分かった。このことから、クラス内変化を捉えるための情報としては BoF によるヒストグラムよりも、HOG 特徴量の方が高い認識結果を得られたため優秀だといえる。

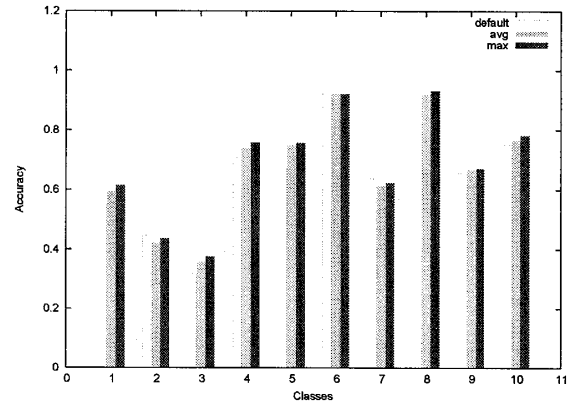


図 4: 実験結果

## 5. まとめと今後の課題

本研究では、一般物体認識におけるクラス内変化を考慮した認識を行った。実験結果では、最高認識率においては一定の効果が見られたものの、安定した精度を実現するには至らなかった。

今後の課題として、クラスタリング手法と背景による影響が挙げられる。HOG 特徴量では背景の影響が非常に大きいため、学習画像における背景との分離は重要となる。また、クラスタリングにおいては、各クラス・データセットにおいて適切なクラスタ数が異なることが実験結果により得られたため、実用にあたりクラスタ数の自動判別が必要となる。

## 参考文献

- [1] 柳井啓司：セマンティックギャップを超えて：人工知能学会誌, Vol. 24, No. 5, pp. 691-699, 2009
- [2] D. G. Lowe: Distinctive image features from scale-invariant keypoints: International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004
- [3] G. Csurka, C.R. Dance, L. Fan, and C. Bray: Visual categorization with bags of keypoint: Proc. of European Conference on Computer Vision, pp. 1-22, 2004
- [4] N. Dalal, B. Triggs: Histograms of oriented gradients for human detection: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 886-893, 2005
- [5] 松島千佳ら：人検出のための Real AdaBoost に基づく HOG 特徴量の効率的な削減法：情報処理学会研究報告, Vol.2009-CVIM-167, No.32