

# Bag-of-Keypoints 表現を用いた Web 画像分類

上 東 太 一<sup>†</sup> 柳 井 啓 司<sup>†</sup>

近年、デジタル画像の爆発的な増大によって、実世界の一般的な画像に対する認識技術への要求が高まっている。そこで、本研究では、新しい認識の手法として G.Csurka らによって提案された bag-of-keypoints 手法を利用して、Web から収集した一般的な画像に対して画像分類実験を行った。Bag-of-keypoints 手法によって画像特徴量を抽出し、SVM によって分類実験を行った結果、100 クラスの分類において分類率 51.1%を実現した。

## Web Image Classification with Bag-of-Keypoints

TAICHI JOUTOU<sup>†</sup> and KEIJI YANAI<sup>†</sup>

Recently, need for generic image recognition is getting larger due to the explosive increase of digital images. Then, we have performed classification experiments for general images gathered from Web employing the bag-of-keypoints method proposed by G.Csurka et al. In the experiments, we have obtained the 51.1% classification rate for 100 object classes.

### 1. はじめに

#### 1.1 背景

近年、デジタルカメラの普及により膨大なデジタル写真が存在するようになった。特に、World Wide Web 上には大量にデジタル写真が存在している。例えば、オンラインアルバムサービスの Flickr<sup>1)</sup> は、個人が撮ったデジタル写真をアップロードすることにより、簡単に Web 上で写真を公開し、多くのユーザーと写真を共有することができるため広く使われている。このような Web に大量にあるデジタル写真は実世界のシーンの画像である。現在では、計算機が画像の意味を理解できないため、このような画像を分類、検索することは人手によって付けられたテキストによるタグによって行われることが一般的で、人手なしには困難なことである。

こうした問題を解決するために一般物体認識の研究が行われている。一般物体認識の研究とは、計算機に画像の意味内容を理解させるための研究である。従来の画像認識の研究で対象としてきた特定の制約の下で撮影された画像とは大きく異なり、制約のない実世界のシーンの画像に対して物体を計算機に認識させる。一般物体認識においては、多様で質の高い学習画像となり得る一般画像データベースが重要である。

実世界シーンの画像においては、同一種類でも様々な画像が存在する。例えば、正面を向いているライオ

ンもいれば、横を向いたり、寝そべっていたり、走っていたり、また動物園の檻の中にライオンもいる。しかし、画像データベースのデファクトスタンダードである Corel 社の Corel Image Gallery は元々画像認識のためのデータセットではないため、対象によっては認識の難易度が大きく異なる問題点や、画像同士の見た目が良く似ていて、研究用のデータベースには不適切であるという問題点がある。よって、同一コンセプト内の、これらの多様な画像を集めてデータベース化することが一般物体認識（つまり様々なライオンを認識する）のためには重要になってくる。すなわち、一般画像データベースを作ることが一般物体認識の研究にとっては重要である。

#### 1.2 目的

本研究の目的は、最終的には一般画像データベースの構築することであるが、今回の発表では、データベースの構築に必要な画像選別方法として新しい画像認識の手法を導入し、分類実験を行うことによって、その有効性を検証することを目的とする。新しい画像認識の手法として、G. Csurka らによって提案された bag-of-keypoints 手法<sup>2)</sup>を応用する。具体的には、当研究室のシステムで Web から集めてきた画像から特徴量を抽出し、その特徴量を用いてカテゴリー分類実験を行う。

#### 1.3 研究の概要

本研究では、画像データをその画像のコンセプトにあったクラスに分類するシステムを作成し、一般画像の分類実験を行う。今回作成したシステムは、画像の部分的な特徴を用いて、それらから画像を認識させる

<sup>†</sup> 電気通信大学大学院 電気通信学研究科 情報工学専攻  
Department of Computer Science, The University of  
Electro-Communications

システムである。特徴量として局所特徴を用いる場合、画像を 1 つの特徴ベクトルに表現するためにテキスト検索の bag-of-words 手法を画像認識に応用させた bag-of-keypoints と呼ばれる手法を用いて、画像を局所特徴の出現頻度で表現する。本実験では、画像データセットには、当研究室で Web から自動収集した画像を使用した。この画像を分類することにより bag-of-keypoints 手法を評価する。分類器には、Nearest Neighbor と SVM を使用した。

まず、1 つめの実験は、特徴量としてカラーヒストグラム、分類器は Nearest Neighbor を用いたシステムと、特徴量として SIFT 特徴、分類器は前記のシステムと同様に Nearest Neighbor を用いたシステムを比較した。

2 つめの実験では、bag-of-keypoints のシステムの分類器を Nearest Neighbor から SVM に変更してシステムの性能をみた。

この 2 つの実験では、10 種類の画像のデータセットを用いてマルチクラス分類を行い、さらには 100 種類の画像データセットで分類実験を行った。

10 種類の分類実験では 84% の分類率、100 種類では 51% の分類率を得られた。

## 2. 関連研究

本研究では、画像特徴に、物体の形状特徴を表現できる SIFT 特徴を用いている。SIFT 特徴は、David Lowe によって考案され<sup>3)</sup> た局所特徴である。局所特徴には、同一の物体であれば違う画像でも同じ場所が特徴点として検出される特性がある。この特性を利用して複数の写真から一枚のパノラマ写真に合成する研究<sup>4)</sup> もされている。SIFT 特徴を用いた研究として、Sivic, J. ら<sup>5)</sup> は映画のフレームを物体認識により検索する研究を行った。この研究は、<sup>4)</sup> のように局所特徴の特性を生かし、特定の物体が映っているフレームを検索できるようにした。しかし、この研究は物体認識なのだが、特定の同一物体に対しての認識であり、物体の種類を区別、分類する認識ではない。

G. Csurka ら<sup>2)</sup> は、初めて bag-of-keypoints で一般物体認識を行った。この研究でも特徴量は SIFT 特徴を使用していて、分類器に Naive Bayes と SVM を用いたシステムで 7 クラスのデータベースで分類実験をした。

これらの研究は、SIFT 特徴と bag-of-keypoints モデルを用い、方法としてはシンプルであるにも関わらず、従来に比べて高い性能を示している。そこで、本研究では、Sivic, J. ら、G. Csurka らの研究を取り入れて、SIFT 特徴と bag-of-keypoints モデルを用いた物体の種類を分類する画像認識システムの実現を行う。

## 3. 画像認識の方針

本研究では、Web から自動収集した画像を画像データセットとして使用し、これらの画像データをその画

像のコンセプトにあったクラスに分類するシステムの作成を試みる。

画像認識の流れとしては、まず、学習データ用とテストデータ用の画像を用意し、全画像から特徴を抽出する。次に、学習データの特徴を用いて分類器を学習させる。最後に、その分類器によってテストデータのクラスを決定する。

本実験では、画像データセットには当研究室で Web から自動収集した画像を使用する。特徴量には、局所特徴の SIFT 特徴を使用する。特徴量として局所特徴を用いる場合、画像を 1 つの特徴ベクトルに表現するために bag-of-keypoints 手法<sup>2)</sup> (図 1) を用いて画像を局所特徴の出現頻度で表現する。実際の処理としては、局所特徴の特徴ベクトルを大量に集め、それらを k-means でベクトル量子化することにより、代表の特徴ベクトルを求める。このベクトル量子化された特徴は visual words と呼ばれ、これらを記述子として用いて、再び画像の局所特徴を 1 つの特徴ベクトルで表す。分類器には、Nearest Neighbor と SVM を使用する。まず、1 つめの実験は、特徴量としてカラーヒストグラムを、分類器は Nearest Neighbor を用いたシステムと、特徴量として SIFT 特徴を、分類器は前記のシステムと同様に Nearest Neighbor を用いたシステムを比較する。

2 つめの実験では、bag-of-keypoints のシステムの分類器を Nearest Neighbor から SVM に変更してシステムの性能をみる。

これらの手法を用いて 10 種類及び 100 種類各種類 50 枚の正解画像データセットを分類するマルチクラス分類実験を行う。

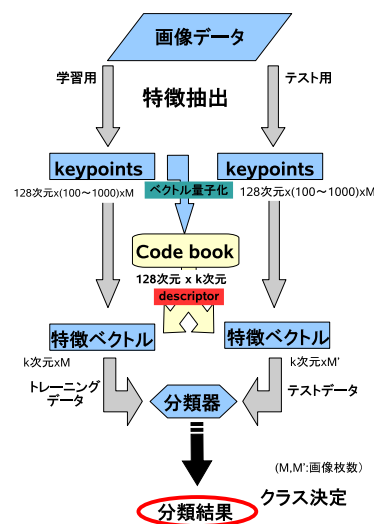


図 1 bag-of-keypoints 手法

## 4. 画像認識方法

### 4.1 特徴抽出

#### 4.1.1 カラーヒストグラム

カラーヒストグラムは最も単純で、類似画像検索によく用いられている手法である。色を数値化するために通常 3 つの軸を持つ 3 次元空間で表され、画像の色もこの色空間で表されている。カラーヒストグラムは、遮蔽や視点変化に安定な物体の表現であり、多数の物体の区別が可能である。しかし、物体の形状は表現することができない。したがって、色の分布が似ていても、画像の内容的には全く異なった画像を同一クラスだと分類してしまうこともある。本実験ではこのカラーヒストグラムを画像の特徴量として抽出する。

#### 4.1.2 局所特徴 (SIFT)

SIFT(Scale Invariant Feature Transformation)<sup>3)</sup> は David Lowe によって考案された。SIFT アルゴリズムは特徴点検出部と特徴点周辺の特徴量記述部に分けられる。特徴点検出部では、DOG(Difference of Gaussian) により、原画像を数段階に縮小したマルチスケール画像群から濃淡変化の大きな特徴的な点を多数検出し、特徴量記述部において、その特徴点周辺の局所領域の特徴量を 128 次元特徴ベクトルで表現する。この特徴は同じ対象であれば異なる画像でも、同じ場所が特徴点として検出される。学習、テスト全画像に対して局所特徴を 1 枚の画像につき約 100~1000 個検出し、SIFT 特徴で表現する。SIFT 特徴の概要は以下のようになっている。

- ガウシアン差分を最大にすることによりスケールを決定
- 主要な勾配方向として局所方向を決定  
(すべての更なる計算をスケールと回転に対し不変にするために、このスケールと方向を使う)
- 勾配方向ヒストグラムを計算し、128 次元ベクトル ( $4 \times 4 \times 8$ ) で表現
- 強度変化に強くするために記述子を正規化

この特徴量は、アフィン変換に対して耐性のある類似性の検出が行えるようになるとされている。つまり、拡大、縮小、回転、歪み、遮蔽、3D 視点の変化に強く、ノイズが入った画像に安定して類似性が検出される。

SIFT のプログラムは C++ 言語で実装されているフリーソフト SIFT++<sup>6)</sup> を利用する。この機能として、入力としてグレースケールの pgm ファイルを与えると、その画像の特徴点抽出とその特徴点に関する座標、スケール、方向、SIFT 特徴量 (128 次元ベクトル) を計算してくれる。なお、SIFT++ はすべて C++ 言語で書かれているので高速である。

### 4.2 Bag-of-keypoints モデル

Bag-of-keypoints モデル<sup>2)</sup> とは、画像を局所特徴の集合と捉えた手法である。局所特徴の特徴ベクトルをベクトル量子化し、visual words と呼ばれる特徴ベクトルを生成する。それらをまとめたものを code book

と呼び、それを記述子として画像の特徴ベクトルを生成する。よって、画像は visual word の集合 (bag) として表現されるわけである。Bag-of-keypoints を用いた画像認識の流れは以下になる。

- (1) 全画像データから特徴を抽出
- (2) それをベクトル量子化し、codebook を作成
- (3) codebook をもとに学習画像の特徴ベクトルを生成
- (4) 同様にテスト画像の特徴ベクトルも生成し、分類器により画像がどのカテゴリに属するか決定

#### 4.2.1 Visual words の生成

まず、visual word を生成するために SIFT 特徴を使い全画像から局所特徴を検出する。次に、各画像の局所特徴が抽出された画像スケール (4.1.2 参照) が大きいものから 100 個だけ選択し、それらをベクトル量子化することにより、似たような局所特徴のクラスタができる。このとき、いくつのクラスタでクラスタリングするかで code book の大きさが決まる。最後に、それぞれクラスタの中心を計算することにより代表的な特徴ベクトルを求め、それらが visual words (図 2) となる。

本来、visual words は学習画像からのみ生成されるべきなのだが、本研究では、評価方法の都合により学習画像の組み合わせが何通りもあるため、すべての画像データセットから visual words を生成する。

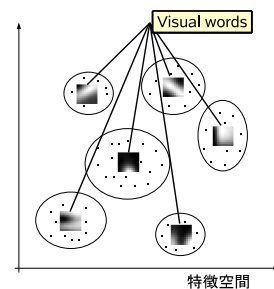


図 2 ベクトル量子化による visual words の生成

##### 4.2.1.1 ベクトル量子化

本実験では、ベクトル量子化は最も単純でよく使われている k-means によってクラスターリングする。k-means では予めクラスタ数 k の値を決める必要があり、k の数によって code book の大きさが決まる。本実験では、k の値は、100, 300, 500, 800, 1000 の 5 つのサイズでベクトル量子化を行う。

##### 4.2.2 Code book による画像の特徴ベクトル表現

ここでは、作成した code book から画像の特徴ベクトル (bag-of-keypoints) を生成する方法を説明する。

画像は code book に基づいたコードパターン出現頻度 (ヒストグラム) によって表現される (図 3)。画像から抽出された局所特徴は code book 中の最も近い visual word に割り当てられる。特徴ベクトルは code

book の大きさ  $k$  と等しい  $k$  次元のヒストグラムとして表現される。しかし、抽出される局所特徴の数は画像によって変わってくる。よって、出現頻度を特徴の数で正規化する。

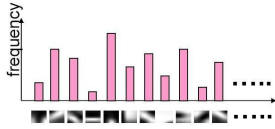


図 3 画像から抽出された局所特徴の出現頻度

#### 4.3 分類・識別

画像から特徴ベクトルを抽出したら、学習画像とテスト画像の特徴ベクトルを比較して学習画像がどのクラスに属するか決定する。

画像から抽出した特徴ベクトルを分類するには、そのクラスかどうかを分類する 1/0 の場合と、どのクラスに分類するかというマルチクラスの場合がある。本研究では、後者のマルチクラスで分類している。

その分類に用いる値として、画像から抽出した特徴量から学習画像との類似度を算出する。この算出方法としてヒストグラムインターセクションを用いた最近傍分類法と SVM<sup>multiclass</sup> (Multi-Class Support Vector Machine)<sup>7)</sup> を用いる。

##### 4.3.1 ヒストグラムインターセクションを用いた最近傍分類法

###### 4.3.1.1 ヒストグラムインターセクション

これは M.J.Swain によって提案された<sup>8)</sup>。  $I_j$  をテスト画像の  $j$  ビン目のヒストグラム、  $M_j$  を学習画像の  $j$  ビン目のヒストグラムとすると、ヒストグラムインターセクション  $S_{IM}$  は、

$$S_{IM} = \sum_{j=1}^n \min(I_j, M_j)$$

と計算され、これを類似度とする。

###### 4.3.1.2 最近傍分類 (Nearest Neighbor)

統計的パターン認識法の 1 つで、テストデータに最も近い学習データを探して (最近傍探索)、その学習データのクラスに分類する方法である。最も単純な分類器ではあるが、どのようなデータ分布であっても平均的な性能を発揮する。

###### 4.3.2 SVM (Support Vector Machine)

SVM<sup>9)</sup> はニューロンのモデルとして最も単純な線形しきい素子を用いて、2 クラスのパターン識別器を構成する手法である。1960 年代に Vapnik らが考案した Optimal Separating Hyperplane を起源とし、カーネル学習法と組み合わせると非線形の識別器になる。この拡張はカーネルトリックと呼ばれる手法で、このカーネルトリックのおかげで、現在知られている多くの手法の中でも最も認識性能の優れた学習モデルの一つであると考えられている。この SVM は基本的には

2 つのクラスの識別しか対応していない、マルチクラスで識別する場合、複数の SVM を組み合わせるなどの工夫が必要であったが、マルチクラス識別の学習プログラムとして、SVM<sup>light</sup><sup>10)</sup> のマルチクラス対応の識別器である、SVM<sup>multiclass</sup> (Multi-Class support Vector Machine)<sup>7)</sup> がある。これは、内部で SVM を組み合わせて、出力はマルチクラスで識別してくれる。本研究ではこれを用いる。

##### 4.3.2.1 カーネル関数

線形識別器は 2 クラスが線形分離可能であるときには高い認識率を期待できるが、非線形で複雑な問題に対してはその限りではない。そこで、非線形な写像  $\Phi$  で写像される先での内積 ( $\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}')$ ) は、元の空間で定義されるカーネル関数  $K(\mathbf{x}_1, \mathbf{x}_2)$  の値と一致するものとする。本実験では、線形識別関数の linear と、カーネル関数の非線形識別関数の radial basis function を用いる。

- linear (線形関数)  
 $y = \text{sign}(\mathbf{w}^T \mathbf{x} - h)$
- radial basis function (非線形関数)  
 $K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{2\sigma^2}\right)$

非線形関数の radial basis function はガウシアン関数であり、これは SVM 上では、Radial Basis Function (RBF) ネットワークと同じ構造になり、構造的には従来のニューラルネットワークと同様である。

## 5. 実 験

### 5.1 実験データセット

実験データとして、1000 種類各 1000 枚の一般画像データベース (すべて正解画像) を目標として当研究室のシステムによって Web から自動収集されている画像を使用した。

その画像データベースのうち、10 種類 (表 1)、及び 100 種類 (表 2) 選んで (10 種類で選んだものはすべて 100 種類に含んでいる)、各 50 枚の人手で正解画像とした画像を用いた。これらの画像データは、学習データ、評価データとして利用する。100 種類は、ジャンル分けしたものを表 2 で示す。

データセット 1) 10 種類 (表 1, 図 4)

「田植え」、「日の出」は風景のジャンルとした。「田植え」は、人が田んぼに稲を植えている画像である。「日の出」の画像は、ほとんどが山か海から朝日が上る風景のものである。「ネジ」、「靴」、「レール」、「グローブ」、「標識」は人工物で、「ネジ」には一個から数個のネジが写っていて、背景は無い。「靴」は、女性もののハイヒールや、革靴、運動靴などの画像がある。これも「ネジ」と同様背景はないが、靴自体に様々な色がある。「レール」には、真っ直ぐのびているものから、カーブしているもの、分岐点のものなど様々ある。「グローブ」は、野球グローブやスキー手袋などの画像である。「標識」は、道路標識が一番多い。スクール

ゾーンや合流地点や分岐地点を表す標識、道路の情報の看板などである。「馬」、「カニ」、「寿司」は動物、水生物、食べ物のジャンルとした。「馬」の画像には、馬だけが写っているのものや、競馬の馬、ジョッキーが乗っている馬など、馬以外にも多くのものが写り込んでいる。「カニ」は、生け簀に入っているカニの画像もあるが、ほとんどの画像が赤で統一されている。「寿司」は、ちらし寿司ではなく、にぎり寿司だけの画像である。これらの画像を各種 50 枚ずつ計 500 枚のデータセットを用意した。

#### データセット 2) 100 種類 (表 2)

予め、表 2 に示すような風景、自然物、人工物、スポーツという大まかなジャンルを決め、さらに、それぞれサブジャンルを決めて分割していった。風景のジャンルには 18 種類 900 枚、自然のジャンルには 26 種類 1300 枚、人工物のジャンルには 52 種類 2600 枚、スポーツのジャンルには 4 種類 200 枚、合計 100 種類 5000 枚のデータセットを用意した。

表 1 10 種類の内容

| クラス名 (種類名) | ジャンル |
|------------|------|
| 田植え        | 風景   |
| 日の出        | 風景   |
| ネジ         | 人工物  |
| 靴          | 人工物  |
| レール        | 人工物  |
| グローブ       | 人工物  |
| 標識         | 人工物  |
| 馬          | 動物   |
| カニ         | 水生物  |
| 寿司         | 食べ物  |

表 2 100 種類の内容

| ジャンル | サブジャンル | クラス名 (種類名)                                            |
|------|--------|-------------------------------------------------------|
| 風景   | 行事     | 花火大会、ねぶた、フリーマーケット、雪祭り、阿波踊り、田植え                        |
|      | 場所     | 紅葉、滝、海岸、湖、温泉、港、鍾乳洞、噴水、流水、川、遊園地、日の出                    |
| 自然物  | 動物     | 猫、犬、パンダ、ウサギ、トカゲ、カンガルー、コアラ、ヘビ、馬                        |
|      | 水生物    | いくら、はたて、イルカ、うなぎ、まぐろ、金魚、カニ                             |
|      | 植物     | 紫陽花、ツツジ、バラ、ヒマワリ、チューリップ、マングローブ、アスバラガス、にんじん、ぶどう、しいたけ    |
| 人工物  | 乗物     | バス、自転車、オートバイ、電車、新幹線、気球、ヘリコプター、カヌー                     |
|      | 建造物    | 橋、教会、ビル、石垣、天守閣、ダム、灯台、神社、寺院、トンネル、レール、標識                |
|      | 小物     | 財布、切手、けし、指輪、ネックレス、リュック、マフラー、ストロー、スーツ、くつ、グローブ、ネジ       |
|      | 食物     | ラーメン、ケーキ、ピザ、かりんとう、いちご、餃子、みかん、アイスクリーム、スバゲッティ、ハンバーガー、寿司 |
|      | 楽器     | ピアノ、トランペット、バイオリン、ハープ、太鼓                               |
|      | 固有名詞   | モアイ、こいのぼり、議事堂、アンコールワット                                |
| スポーツ |        | ロッククライミング、卓球、剣道、相撲                                    |

## 5.2 実験方法

ここでの目的は、一般画像データベースからの 10 種類 (図 4) と 100 種類の実験データセットをマルチ

クラス分類することである。その方法は、

- 10 種類の分類実験

**手法 1** カラーヒストグラム + Nearest Neighbor

**手法 2** bag-of-keypoints + Nearest Neighbor

**手法 3** bag-of-keypoints + SVM(線形カーネル linear)

**手法 4** bag-of-keypoints + SVM(非線形カーネル radial basis function)

- 100 種類の分類実験

- 手法は 10 種類実験の性能のよかった手法を用いる

で行う。10 種類での bag-of-keypoints 手法の実験は、code book のサイズを 100, 300, 500, 800, 1000 で行い、その中で最も良かったもので 100 種類の分類実験を行うことにする。

### 5.3 評価方法

分類実験の手法は、実験データセットの 5-fold cross validation によって評価する。各種別の画像セットを 5 つのグループに分割し、ある 1 つのグループを評価データとして利用するとき、残りの 4 グループは学習データとして分割器を訓練させる。これを 5 通りそれぞれのグループが評価データとして分類実験を行い、その 5 回の結果の平均を分類結果とする。

分類結果の評価に用いる基準として、正確性の観点からみた適合率 ((正しく分類されたデータ集合)/(分類されたデータ集合)) と、完全性の観点からみた再現率 ((正しく分類されたデータ集合)/(分類されるべきデータ集合))、総合的な尺度として F-値 (F-measure)、さらに、与えられた画像がどのクラスに分類されたか知るために混行列 (confusion matrix) で表してみる。

#### 5.3.1 F-値 (F-measure)

F-値は再現率 (recall) と適合率 (precision) の双方の値が大きいときに、大きい値をとる。これは再現率と適合率の調和平均 (逆数の平均値の逆数) で計算される。

$$F = \frac{2}{\frac{1}{recall} + \frac{1}{precision}}$$

この値は 0 ~ 1 の範囲をとり、値が大きいほどよい。

#### 5.3.2 混行列 (confusion matrix)

マルチクラス分類システムを評価する基準として混行列  $M_{ij}$  をこのように定義する。

$$M_{ij} = \frac{|\{I_k \in C_j : h(I_k) = i\}|}{|C_j|}$$

ここで、 $i, j \in \{1, \dots, N_c\}$  ( $N_c$  はクラスの数)、 $C_j$  はクラス  $j$  のテストデータ集合、 $h(I_k)$  は、画像  $I_k$  に対して分類器から出力された値の最大値を得たクラスとする。

## 5.4 実験結果

### 5.4.1 10 種類分類実験の結果

図 4 のような 10 種類のクラスで分類実験を行った結果について、手法 1 と手法 2 の結果を再現率で表したものを表 3 で示し、適合率で表したものを表 4 で示





図 4 10 種類の画像のサンプル (左上から、田植え、ネジ、レール、馬、寿司、日の出、靴、グローブ、カニ、標識)

す。手法 2 では code book のサイズ別に表記した。

次に、手法 2 の分類器を Nearest Neighbor から SVM に変えて分類実験した結果を表 5 と表 6 で示す。SVM はカーネルを 2 通り試したため、カーネル別に分けてから code book サイズで分けて表記した。

また、クラス別にそれぞれの手法の F-値を計算したものを表 5 で示す。ただし、bag-of-keypoints を用いた手法においては、結果がよかった code book サイズの実験結果のみを示す。

最後に、10 種類の F-値の平均が最も良かった、カーネルに radial basis function を用いた bag-of-keypoints+SVM の手法 (手法 4) で code book サイズを 800 としたときの結果を混合行列にして表 7 で示す。

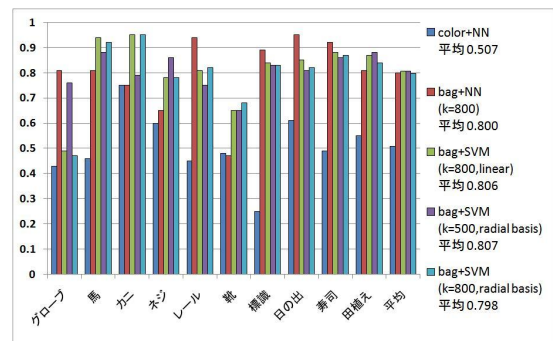


図 5 10 種類の F-値とその平均

#### 5.4.2 100 種類分類実験の結果

10 種類分類の結果より、SVM を用いた手法においてカーネルの違いによる大きな結果の違いがなかったため、学習時間が圧倒的に短い線形 (linear) カーネルによる SVM を 100 種類分類において採用する。

100 種類分類の結果を 100 種類の平均再現率と平均適合率を用いて表 8 で示す。さらに、F-値で評価した結果の上位 10 種類と、下位 10 種類を表 9 で示す。結果が特に良かったもの、悪かったものを図 6 で示す。

表 8 100 種類の分類結果

| 平均再現率 | 平均適合率 |
|-------|-------|
| 0.511 | 0.489 |

表 3 カラー+NN と bag-of-keypoints+NN の再現率

| クラス名 | カラー<br>ヒストグラム | bag-of-keypoints(code book size(c)) |             |             |             |             |
|------|---------------|-------------------------------------|-------------|-------------|-------------|-------------|
|      |               | 100                                 | 300         | 500         | 800         | 1000        |
| グローブ | 0.40          | 0.74                                | 0.92        | 0.86        | 0.78        | <b>0.96</b> |
| 馬    | 0.42          | <b>0.86</b>                         | 0.74        | 0.82        | 0.72        | 0.80        |
| カニ   | 0.68          | 0.90                                | 0.90        | <b>0.98</b> | 0.96        | <b>0.98</b> |
| ネジ   | 0.68          | 0.54                                | <b>0.74</b> | 0.66        | 0.26        | 0.56        |
| レール  | 0.60          | 0.92                                | 0.86        | 0.92        | <b>1.00</b> | 0.94        |
| 靴    | 0.48          | 0.34                                | 0.34        | <b>0.38</b> | 0.22        | 0.34        |
| 標識   | 0.18          | 0.60                                | 0.82        | 0.70        | 0.54        | <b>0.90</b> |
| 日の出  | 0.60          | 0.70                                | 0.58        | 0.64        | 0.62        | <b>0.90</b> |
| 寿司   | 0.48          | 0.84                                | <b>0.92</b> | 0.88        | 0.90        | 0.90        |
| 田植え  | 0.52          | 0.94                                | <b>0.98</b> | 0.88        | 0.92        | 0.82        |
| 平均   | 0.52          | 0.74                                | 0.78        | 0.77        | 0.69        | <b>0.81</b> |

表 4 カラー+NN と bag-of-keypoints+NN の適合率

| クラス名 | カラー<br>ヒストグラム | bag-of-keypoints(code book size(c)) |             |             |             |             |
|------|---------------|-------------------------------------|-------------|-------------|-------------|-------------|
|      |               | 100                                 | 300         | 500         | 800         | 1000        |
| グローブ | 0.46          | 0.59                                | 0.61        | <b>0.72</b> | 0.46        | 0.70        |
| 馬    | 0.52          | 0.76                                | 0.69        | <b>0.87</b> | 0.79        | 0.83        |
| カニ   | 0.83          | 0.75                                | <b>0.82</b> | 0.66        | 0.70        | 0.61        |
| ネジ   | 0.54          | <b>0.96</b>                         | 0.78        | 0.80        | 0.61        | 0.78        |
| レール  | 0.36          | 0.86                                | 0.75        | 0.87        | 0.87        | <b>0.94</b> |
| 靴    | 0.48          | 0.43                                | 0.74        | 0.68        | 0.35        | <b>0.78</b> |
| 標識   | 0.43          | 0.76                                | <b>1.00</b> | 0.75        | 0.62        | 0.89        |
| 日の出  | 0.62          | 0.84                                | 0.80        | 0.98        | <b>1.00</b> | <b>1.00</b> |
| 寿司   | 0.51          | 0.87                                | 0.90        | 0.82        | 0.77        | <b>0.94</b> |
| 田植え  | 0.58          | <b>0.80</b>                         | 0.79        | 0.79        | 0.77        | <b>0.80</b> |
| 平均   | 0.53          | 0.76                                | 0.79        | 0.79        | 0.69        | <b>0.83</b> |

表 9 100 種類の F-値の TOP10 と WORST10

| クラス名   | TOP10 | クラス名   | WORST10 |
|--------|-------|--------|---------|
| 自転車    | 0.87  | 橋      | 0.03    |
| ネックレス  | 0.78  | 港      | 0.05    |
| 阿波踊り   | 0.77  | 神社     | 0.05    |
| 花火大会   | 0.76  | 猫      | 0.08    |
| スパゲッティ | 0.76  | 相撲     | 0.11    |
| 温泉     | 0.76  | アスパラガス | 0.14    |
| 剣道     | 0.76  | 川      | 0.17    |
| ねぶた    | 0.75  | うなぎ    | 0.18    |
| 石垣     | 0.74  | トカゲ    | 0.18    |
| スーツ    | 0.73  | ツツジ    | 0.19    |

表 5 bag-of-keypoints+SVM での再現率

| クラス名 | カーネル           |             |             |             |             |                       |             |             |             |             |
|------|----------------|-------------|-------------|-------------|-------------|-----------------------|-------------|-------------|-------------|-------------|
|      | linear         |             |             |             |             | radial basis function |             |             |             |             |
|      | code book size |             |             |             |             | code book size        |             |             |             |             |
|      | 100            | 300         | 500         | 800         | 1000        | 100                   | 300         | 500         | 800         | 1000        |
| グローブ | 0.72           | 0.62        | 0.6         | 0.34        | <b>0.92</b> | 0.80                  | 0.60        | 0.64        | 0.34        | <b>0.92</b> |
| 馬    | 0.80           | 0.56        | 0.80        | 0.90        | 0.96        | 0.74                  | 0.56        | 0.80        | 0.86        | <b>0.98</b> |
| カニ   | 0.50           | 0.62        | 0.68        | <b>0.94</b> | 0.46        | 0.58                  | 0.66        | 0.72        | <b>0.94</b> | 0.48        |
| ネジ   | 0.90           | <b>0.94</b> | 0.80        | 0.82        | 0.88        | 0.80                  | <b>0.94</b> | 0.84        | 0.86        | 0.84        |
| レール  | 0.96           | 0.94        | 0.96        | <b>1.00</b> | <b>1.00</b> | 0.98                  | 0.94        | 0.96        | <b>1.00</b> | <b>1.00</b> |
| 靴    | 0.00           | 0.10        | 0.44        | 0.54        | 0.04        | 0.06                  | 0.10        | 0.52        | <b>0.56</b> | 0.12        |
| 標識   | 0.76           | 0.76        | <b>0.96</b> | 0.84        | 0.90        | 0.90                  | 0.78        | <b>0.96</b> | 0.84        | 0.90        |
| 日の出  | 0.62           | 0.88        | 0.96        | <b>1.00</b> | 0.82        | 0.60                  | 0.82        | 0.92        | 0.98        | 0.80        |
| 寿司   | 0.92           | 0.80        | 0.83        | 0.90        | <b>0.98</b> | 0.92                  | 0.84        | 0.84        | 0.90        | <b>0.98</b> |
| 田植え  | 0.90           | 0.62        | 0.90        | 0.84        | 0.94        | 0.90                  | 0.64        | 0.90        | 0.82        | <b>0.96</b> |
| 平均   | 0.71           | 0.68        | 0.79        | <b>0.81</b> | 0.79        | 0.73                  | 0.69        | <b>0.81</b> | <b>0.81</b> | 0.80        |

表 6 bag-of-keypoints+SVM での適合率

| クラス名 | カーネル           |             |             |             |             |                       |             |             |             |             |
|------|----------------|-------------|-------------|-------------|-------------|-----------------------|-------------|-------------|-------------|-------------|
|      | linear         |             |             |             |             | radial basis function |             |             |             |             |
|      | code book size |             |             |             |             | code book size        |             |             |             |             |
|      | 100            | 300         | 500         | 800         | 1000        | 100                   | 300         | 500         | 800         | 1000        |
| グローブ | 0.59           | 0.63        | <b>0.97</b> | 0.85        | 0.88        | 0.56                  | 0.55        | 0.94        | 0.74        | 0.85        |
| 馬    | 0.89           | 0.70        | 0.98        | 0.98        | <b>1.00</b> | 0.90                  | 0.68        | 0.98        | 0.98        | <b>1.00</b> |
| カニ   | <b>1.00</b>    | 0.66        | 0.89        | 0.96        | <b>1.00</b> | 0.97                  | 0.67        | 0.88        | 0.96        | <b>1.00</b> |
| ネジ   | 0.69           | 0.68        | 0.87        | 0.75        | 0.72        | 0.74                  | 0.71        | <b>0.88</b> | 0.72        | 0.71        |
| レール  | 0.60           | 0.65        | 0.59        | 0.68        | 0.63        | <b>0.77</b>           | 0.64        | 0.62        | 0.70        | 0.65        |
| 靴    | 0.00           | 0.23        | 0.81        | 0.82        | 0.20        | 0.18                  | 0.24        | 0.87        | <b>0.88</b> | 0.35        |
| 標識   | 0.90           | <b>0.93</b> | 0.73        | 0.84        | 0.83        | 0.82                  | <b>0.93</b> | 0.73        | 0.82        | 0.79        |
| 日の出  | 0.70           | 0.68        | 0.67        | 0.68        | 0.69        | <b>0.81</b>           | 0.75        | 0.73        | 0.71        | 0.78        |
| 寿司   | 0.84           | 0.70        | 0.91        | 0.87        | <b>0.94</b> | 0.81                  | 0.72        | 0.89        | 0.85        | <b>0.94</b> |
| 田植え  | 0.63           | 0.82        | 0.85        | <b>0.91</b> | 0.77        | 0.61                  | 0.82        | <b>0.87</b> | 0.87        | 0.80        |
| 平均   | 0.68           | 0.61        | 0.74        | 0.83        | 0.77        | 0.72                  | 0.67        | <b>0.84</b> | <b>0.84</b> | 0.79        |

表 7 手法 4(code book size=800) の混同行列

| True class ↓ | グローブ      | 馬         | カニ        | ネジ        | レール       | 靴         | 標識        | 日の出       | 寿司        | 田植え       |
|--------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| グローブ         | <b>17</b> | 0         | 0         | 14        | 7         | 2         | 6         | 1         | 3         | 0         |
| 馬            | 0         | <b>43</b> | 0         | 0         | 1         | 1         | 3         | 0         | 2         | 0         |
| カニ           | 0         | 0         | <b>47</b> | 0         | 1         | 0         | 0         | 0         | 0         | 2         |
| ネジ           | 1         | 0         | 0         | <b>43</b> | 5         | 1         | 0         | 0         | 0         | 0         |
| レール          | 0         | 0         | 0         | 0         | <b>50</b> | 0         | 0         | 0         | 0         | 0         |
| 靴            | 5         | 0         | 0         | 1         | 0         | <b>28</b> | 0         | 12        | 1         | 3         |
| 標識           | 0         | 0         | 0         | 0         | 8         | 0         | <b>42</b> | 0         | 0         | 0         |
| 日の出          | 0         | 0         | 0         | 0         | 0         | 0         | 0         | <b>49</b> | 0         | 1         |
| 寿司           | 0         | 0         | 0         | 0         | 0         | 0         | 0         | 5         | <b>45</b> | 0         |
| 田植え          | 0         | 1         | 2         | 2         | 0         | 0         | 0         | 2         | 2         | <b>41</b> |

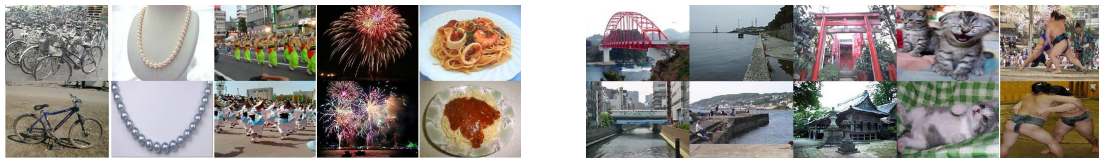


図 6 左図が結果がよかった例, 右図が結果が悪かった例

## 5.5 考 察

### 5.5.1 10 種類の分類

まず, 手法 1 と手法 2 の結果より, カラーヒストグラムと SIFT 特徴 (bag-of-keypoints) の特徴量について考察してみる. 表 3, 表 4 を見ると, 再現率, 適合率とも SIFT 特徴の方が, 圧倒的に認識率が優れている. 特に, 標識の画像は図 4 を見るとわかるように, 標識には, 角や丸みを帯びたものなど他の種類にない物体の形状の特徴を持っている. 逆に, 色の視点から見て

みると, 背景に建物や空が写っていて, 色だけでは判断にくいことがわかる. このことにより, カラーヒストグラムより SIFT 特徴の方が特徴量として有効だと示される.

次に, 手法 2 と手法 3 の結果より Nearest Neighbor と SVM の分類器について考察してみる. 表 3, 表 4 より, 手法 2 の再現率と適合率の平均は 0.76 と 0.77 であり, 手法 3 は表 5, 表 6 より, 線形カーネルを用いた SVM のものは, 0.76 と 0.73 で若干手法 2 の方

が結果が良いが、非線形カーネルを用いた SVM で学習すると 0.77 と 0.77 となり、手法 2 を上回った。

最後に、手法 3 と手法 4 の F-値 (図 5) がほぼ等しいことから、SVM のカーネルの違いがないことがわかる。この結果から、この 10 種類の画像は特徴空間において線形分離可能であったと推測できる。つまり、bag-of-keypoints で表現した特徴ベクトルは、この場合、次元数が 800 次元と極めて高かったために、非線形写像の有無は結果に大きな影響を与えなかった。

### 5.5.2 100 種類の分類

100 種類の分類結果 (表 8) を見てみると、平均再現率、平均適合率とも約 50% の精度が得られた。ランダムサンプリングだと 100 クラスの場合、分類率が 1% であることから、これは十分な分類精度が得られたということになる。最も分類結果が良かった種類は「自転車」で、再現率は 0.94、適合率は 0.81 であった。逆に、最も分類結果が悪かった種類は「橋」で、再現率は 0.02、適合率は 0.04 であった。表 9 をみると、特徴が少ないと思われていた風景のジャンルの画像が予想に反して、上位に 4 種類も入っていた。これは、風景にも局所特徴が有効であることを示している。

## 6. おわりに

本研究では、Web から収集した画像データをその画像のコンセプトにあったクラスに分類するシステムを作成し、一般画像の分類実験を行った。

実験では、10 種類の画像データセットを用いてマルチクラス分類を行い、さらには、100 種類の画像データセットでも分類実験を行った。10 種類の分類実験では、最高で分類率 84% の精度、100 種類でも平均約 50% の分類率を得られた。ベースラインとしてカラーヒストグラムと Nearest Neighbor の手法と比較したが、その差は歴然であった。

一般に、線形分離可能性はデータ数が大きくなればなるほど難しくなり、逆に、特徴空間ベクトルの次元が大きくなるほど易くなる。局所特徴から構築された visual words はいくらでも増やすことができ、それにともない画像を表現する特徴ベクトルの次元数も増える。そうすることによって、特徴空間が広がり、それまでデータの分布が偏っていたり、密集してて、認識しづらかったものが容易に認識できるようになる。しかし、次元数を大きく取りすぎると、認識するにあたってあまり必要でない特徴まで、場合によっては、不利になる特徴まで表現できるようになり、認識率を下げる原因になる恐れがあることがわかった。

今後の課題は、この bag-of-keypoints 手法をデータベースの選別方法として取り入れ、データベースの拡張させることである。さらに、O. Maron<sup>11)</sup> らの研究をもとに、Multiple Instance Learning (MIL) を導入することを考えている。MIL を用いると正解画像から曖昧さ (ノイズ) を差し引いた正解部分の特徴だけを学習することができる。正解部分とは、正解画像と正

解画像の共通部分のことであり、O. Maron らが提案した Diverse Density と呼ばれる新しいフレームワークによって推定される。MIL を用いると正解画像から曖昧さ (ノイズ) を差し引いた正解部分の特徴だけを学習することができる。Chen, Y. ら<sup>12)</sup> は、この MIL と局所特徴を組み合わせることにより、認識対象以外の局所特徴を学習せずに、効率の良い学習をしている。そこで本研究は今後、Chen, Y. らの研究を取り入れ、さらなる認識率の向上を目指す。

## 参 考 文 献

- 1) Flickr: <http://www.flickr.com/>.
- 2) Csurka, G., Bray, C., Dance, C. and Fan, L.: Visual categorization with bags of keypoints, *Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision*, pp.1-22 (2004).
- 3) Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004).
- 4) Brown, M. and Lowe, D.: Recognising panoramas, *Proc. The International Conference on Computer Vision*, pp.1218-1225 (2003).
- 5) Sivic, J. and Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos, *Proceedings of the International Conference on Computer Vision*, Vol.2, pp.1470-1477 (2003).
- 6) Vedaldi, A.: SIFT++, <http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html>.
- 7) Joachims, T.: SVM<sup>multiclass</sup>, [http://www.cs.cornell.edu/People/tj/svm-light/svm\\_multiclass.html](http://www.cs.cornell.edu/People/tj/svm-light/svm_multiclass.html).
- 8) M.J. Swain and D.H. Ballard: Color Indexing, *International Journal of Computer Vision*, Vol.7, No.1, pp.11-32 (1991).
- 9) 栗田多喜夫: サポートベクターマシン入門 (2003).
- 10) Joachims, T.: SVM<sup>light</sup>, <http://www.cs.cornell.edu/People/tj/svm-light/index.html>.
- 11) Maron, O. and Ratan, A.: Multiple-instance learning for natural scene classification, *The Fifteenth International Conference on Machine Learning*, pp.341-349 (1998).
- 12) Chen, Y., Bi, J. and Wang, J.: MILES: Multiple-Instance Learning via Embedded Instance Selection, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 28, No.12, pp.1931-1947 (2006).