

撮影位置の情報を用いた一般画像認識の可能性の検討

八重樫 恵太^{†1} 柳 井 啓 司^{†1}

今日では、Web 上のデジタル写真の普及、およびマッピングサービスの高度化は目覚ましく、それに伴い位置情報付き写真も増加している。一方で、画像認識の分野においてもより高度な手法が提唱されてきた。位置情報を画像の特徴の1つとして用いることにより、画像認識の精度を向上できる可能性が考えられる。本研究では、位置情報と航空写真を対応する画像の特徴量として用いることによって、画像認識の精度向上の可能性を検討する。画像をキーワードごとに分類し、画像と航空写真からそれぞれ SIFT 特徴を抽出し、位置情報とそれらの特徴量を統合したものの組み合わせについて SVM による学習と分類の実験を行う。実験の結果、いくつかのキーワードについて、分類精度の向上が見られた。

A Study on Generic Image Recognition using GPS data

KEITA YAEGASHI^{†1} and KEIJI YANAI^{†1}

Because of recent remarkable spread of digital photos and mapping service on the Web, the number of geotagged photos are increasing greatly. Meanwhile, in the research community of image recognition, research on generic object recognition has progressed greatly for these several years. Then, we propose using geolocation as one of features to recognize geotagged images. In this paper, we examine possibility of improvement on generic image recognition by using geolocation and aerial photos as features in addition to image features extracted from images. We collected geotagged images associated with several keywords from Flickr, and extracted feature vectors by the Bag-of-Keypoints method from each image and each associated aerial photo. Next, we combined feature vectors extracted from images and aerial photos into one vector, and we performed experiments on image classification using SVM. As the result, improvements of the classification rate were obtained for some keywords.

1. はじめに

1.1 背景

今日では、デジタルカメラの普及により、デジタル写真は WWW(World Wide Web) 上に大量に存在する。また、Flickr などの写真共有を行うソーシャルサイトの普及により、個人が多くの写真を収集、整理し、共有することが可能になった。これにより WWW 上のデジタル写真が爆発的に増加することは容易に予想される。一方で、Google Maps^{?)} などのオンラインマッピングサービスも普及し、日々高度化している。利用者は自宅から地図を利用した位置の検索だけでなく、航空写真の閲覧など高度なサービスを利用できるようになった。また、写真の収集に関しても、位置情報を積極的に活用しようとする試みが行われてきた。その例として Flickr^{?)} では、写真に位置情報を付加して投稿し、地図 (Flickr Map) を用いて写真を検索、整理できるようになっている。

前述のように WWW 上には大量の画像が存在するが、現状では計算機を用いて自動で画像を分類、検索することは未だに困難な課題である。こうした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試みとして、一般画像認識^{?)} の試みが広く行われている。

画像認識の精度を向上させるためには、画像から適切な特徴量を抽出する必要があるが、特に近年において SIFT 特徴による局所特徴抽出の手法と、Bag of Keypoints による学習と分類の手法は、一般画像認識の研究に大いに貢献している。

1.2 目的

画像認識においては、画像のみから特徴量を抽出することに特化している。一方、位置情報を用いて画像の分類を行う研究は広く行われているが、現状においては、位置情報は写真の整理や閲覧への活用にとまわっている。位置情報を画像の特徴の1つとして用いることにより、画像認識の精度を向上できる可能性が考えられる。

本研究では、Web 上から収集した位置情報付きの画像と、位置情報に対応する航空写真を用いて、画像

^{†1} 電気通信大学大学院 電気通信学研究所 情報工学専攻
Department of Computer Science, The University of
Electro-Communications

認識の精度が向上するかを比較検討する。

具体的には、画像と航空写真から特徴量を抽出し、画像の特徴量のみで学習・分類した場合と、画像に対応する位置情報、航空写真の特徴量をそれぞれ統合した場合の学習・分類した場合で比較する。最終的には、今後位置情報が画像認識に対して妥当な指標になりうるかを検討する。

2. 関連研究

位置情報を利用した研究は近年になって盛んに行われるようになった。特に位置情報と画像を用いた研究で、写真を位置情報に基づいて整理、分類する研究が広く行われている。

いずれの研究も地理的な位置情報を用いることによって、整理と分類の精度、あるいはシステムのユーザビリティが向上するという結果を出している。Y. Chen²⁾らは、位置情報をコンテンツのテキストからマイニングすることで推定し、RSSに位置情報をタグ付けすることで、時系列及び位置情報の両観点から、コンテンツを効率的にブラウジングするシステムを考案した。彼らはその研究において、位置情報の推測によるコンテンツナビゲーションのパフォーマンスは、自然言語処理や情報検索における、機械学習の技法と比較して優れていることを主張した。

また、位置情報に基づき写真を分類することで、ユーザビリティを向上させる研究として、S. Ahern²⁾らはWorld Explorerというシステムを説明した。この研究では、本研究と同様にFlickrの写真を用いて、位置情報によりk-Meansでクラスタリングし、それらのタグをTF-IDFでスコアリングすることによって写真を整理している。

その研究においては、ユーザビリティという観点から、客観的な数値評価などではなく、被験者による質的なフィードバックを重視している。

藤田²⁾らの研究では、位置情報と実世界を3次元で関連づける研究を行った。この研究では、地理情報を空間的な情報と捉えて写真を検索するシステムを提唱している。

いずれの研究においても、位置情報は情報を検索する手段として用いられている。本研究では、航空写真を用いて位置情報に画像としての特徴を持たせることによって、分類精度が向上するかどうかを検討する。

3. 位置情報と航空写真

3.1 位置情報

本研究で扱う位置情報は、GPSにより収集される緯度、経度の地理的な位置情報である。本節では、本研究で扱う位置情報付き写真と航空写真について説明する。

3.1.1 Flickrの位置情報つき写真

Flickrは写真を投稿、共有することが出来るソーシャルサイトであり、毎日100万以上の画像が投稿されるといわれている。Flickrでは撮影した写真を投稿する場合、その位置情報を付加させることを強く推奨している。従って、今後Flickrに投稿される位置情報つき写真は増え続けると予想される。本研究では、画像認識に用いる画像の対象としてFlickrの位置情報つき写真を収集し、利用する。

Flickrの投稿画像には、メタデータとして位置情報とその精度を付加させることが出来る。また、Flickr APIでは、位置情報を境界範囲として指定して画像を検索することができる。

3.1.2 位置情報付き写真とExif形式

普及している多くのデジタルカメラが、Exif形式をサポートしている。Exif形式の情報が埋め込まれた写真は、画像ファイル単独で位置情報を保持する。

本研究では、Flickrから収集した写真を扱う。したがって、収集する画像データの全てに対してExifが含まれているとは限らないことを想定している。また、位置情報を特徴量として利用することに主眼を置くため、Exif形式の情報は扱わず、画像と位置情報を独立のものとして対応付ける。

3.2 航空写真

Web上でマッピングサービスが普及したことにより、航空写真を身近に入手し、利用することは容易になった。現在では国土交通省など政府機関をはじめ、GoogleやYahoo!などの検索サービスが航空写真を無料でユーザーに提供するようになった。特に国土交通省では過去の航空写真を閲覧できるサービスも提供している。一方、Google MapsやGoogle Earthでは、国境に関係なく地球上の非常に広範囲にわたる航空写真を閲覧することが出来る。

航空写真は、位置情報(緯度経度)と対応しているので、画像の特徴を持った位置情報と見なせる。したがって、位置情報を記述、識別する客観的な手段に成りうると考えられる。

マッピングサービスでは、通常メルカトル図法を採用している。あるズームレベルと位置情報におけるタイルへの座標は、メルカトル投影によって、位置情報からそのズームレベルでのピクセル座標、ピクセル座標からタイル座標に変換することによって得られる。

4. 画像認識における特徴抽出

本節では、画像から認識に必要な特徴を抽出する方法について述べる。画像の特徴を記述する手法としては、本実験では、特徴抽出のために局所特徴の一種であるSIFT特徴を用いる。また、この局所特徴を簡潔に記述するためにBag of Keypoints手法を用いてデータをベクトル量子化する。

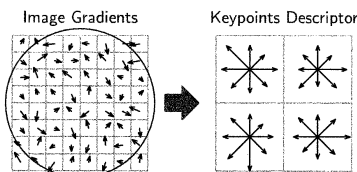


図 1 SIFT 特徴

4.1 SIFT 特徴

SIFT(Scale Invariant Feature Transform)²⁾とは、David Lowe によって提案された特徴点とそれに付随する特徴ベクトルの抽出法である。図 1 に示すように、特徴点周りの局所画像パターンを 128 次元特徴ベクトルで表現する。SIFT 特徴は、画像の拡大縮小、回転や視点のいずれに対してもロバストである。

SIFT 特徴の抽出は、特徴点の抽出とその特徴点における特徴ベクトルの抽出の 2 つのステップに分けることができる。

特徴点抽出の主な方法に、ガウシアン差分 (DoG, Difference of Gaussian) を用いた方法がある。ガウシアン差分は、異なるスケールのガウシアンフィルタで平滑化した画像の差分において、画素の変化しにくい部分を特徴点とする方法である。DoG は、物体画像に関しては精度の高い類似判定に用いることができるが、特徴点の抽出方法はこの限りではない。抽出された特徴点に対して、次の手順で特徴ベクトルを抽出する。

- (1) 勾配方向ヒストグラムを計算し、128 次元ベクトル ($4 \times 4 \times 8$) で表現する
- (2) 強度変化に強くするために記述子を正規化するこの特徴量は、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。

4.2 GRID 点抽出

Web 上に存在する写真は、シーン画像が多く、殆んどの場合特定の物体を明確に捉えているとは限らないので、本研究では、GRID 点抽出で SIFT 特徴量の抽出を試みる。GRID 点抽出では、格子状に点を、SIFT 特徴ベクトル計算のための特徴点として利用する。GRID 点抽出による SIFT 特徴の抽出手順は以下のようになる。

- (1) 格子点の間隔を決定する。
- (2) 画像から格子点を抽出し、予め決められた複数のスケールで、それぞれの点について勾配方向を計算する。
- (3) 抽出した特徴点に対して、SIFT 特徴量を計算する。

4.3 Bag of Keypoints モデル

Bag of Keypoints モデル²⁾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 Codebook 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。これにより、画像を Visual Words の集合 (bag) として表現することができる。Bag of Keypoints の手法の手順を図 2 に示す。また、Bag-of-keypoints の画像認識の流れは次のようになる。

- (1) 全画像データから特徴を抽出する
- (2) それをベクトル量子化し、Codebook を作成する
- (3) Codebook をもとに学習画像の特徴ベクトルを生成する
- (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する

4.3.1 Codebook の生成

まず、Visual Words を生成するために GRID 点における SIFT 特徴を用いて全ての画像から局所特徴を抽出する。次に抽出したもののうち、学習画像についての局所特徴をベクトル量子化し、各々のクラスタの中心を求めることにより Visual Words を求め、Codebook とする。

ベクトル量子化は、最も単純なクラスタリング手法である k-Means 法を用いる。Codebook の大きさは、クラスタ数 k に依存する。本研究の実験では、 $k = 300$ に固定してベクトル量子化を行う^{*1}。k-Means 法における距離尺度として本研究ではユークリッド距離を用いる。

5. 学習と分類

本研究では、学習と分類の手段として SVM(Support Vector Machine) を用いる。

5.1 SVM(Support Vector Machine)

SVM 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。本実験では SVM を実行するツールとして SVM^{light?}を用いる。

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 Φ で写像される先での内積 ($\Phi(x) \cdot \Phi(x')$) は、元の空間で定義されるカーネル関数 $K(x_1, x_2)$ の値と一致するものとする。本実験では、カーネル関数として非線形識別関数の radial basis function(RBF)

*1 相対的な比較が目的であるので、 k は大きいほど実験精度が良くなるが、 k の大きさは処理速度との間でトレードオフの関係となるので、ここではこの値を用いる

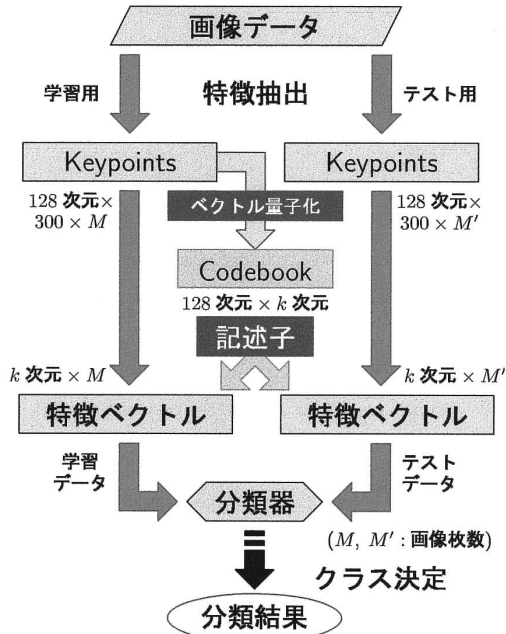


図2 Bag of Keypoints 手法

を用いる。

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\frac{-\|\mathbf{x}_1 - \mathbf{x}_2\|}{2\sigma^2}\right) \quad (1)$$

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。

6. 方法

6.1 方針

画像と航空写真から GRID 点に沿って SIFT 特徴を抽出し、Bag of Keypoints の手法で Codebook を作成する。位置情報 (緯度経度) については、これをそのまま特徴量として用いる。これらの特徴量を適宜統合し、SVM(RBF カーネル) による 2 値分類を用いて学習と分類の実験を行う。

6.2 概要

本研究では、位置情報付き画像に、位置情報を特徴として付加した状態で画像認識の実験を行うことで、画像認識の精度にいかなる変化が起こるかを検証する。本実験で画像に付加する情報として、緯度経度の情報と航空写真を用いる。ただし、位置情報と航空写真については、それぞれ単独での認識についても検証するものとする。

本研究の方法の手順を以下に示す。

- (1) 位置情報付き画像および航空写真を収集する。

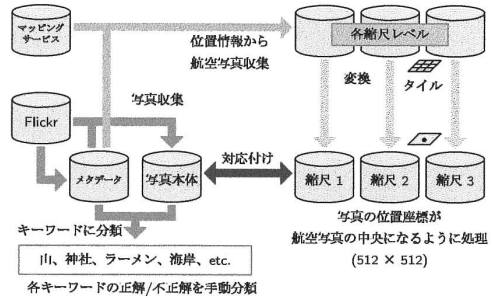


図3 写真の収集とキーワード分類の手順

- (2) 各画像について、前処理と特徴量の抽出を行う。
- (3) 特徴量に基づいた学習データを作成し、分類実験を行う。

6.3 位置情報付き写真の収集と分類

位置情報付き写真は Flickr API により行う。Flickr API では写真の検索を行うことができ、地理座標を境界範囲に指定することができる。

検索結果は写真のメタデータで、XML 形式で返される。このメタデータに基づき写真の URL を生成し、収集を行う。写真の URL の生成方法は Flickr API の詳細と共に公開されている。

収集された写真を、幾つかのキーワードの集合に分類する。例えば、「山が目立つように写っている写真」は、「山」というキーワードに分類する。収集された写真のメタデータには、タグが含まれる。タグによって、写真を分類することができる。本研究では、収集された写真のメタデータから、分類するキーワードと、それに関連する語句で検索し、その語句を含むものをキーワードに分類する。

各キーワードの画像に対して、特徴抽出と学習、分類を行う。

6.4 写真の特徴抽出

写真に対して、10 ピクセルごとに GRID 点抽出を行い、それらの点に基づき SIFT 特徴量を計算する。

航空写真についても、Flickr の画像と同様に SIFT 特徴を抽出する。

6.5 学習データの作成と分類

本節では、収集した画像とメタデータを用いた、学習と分類の方法について説明する。

6.5.1 Codebook の作成

学習データ作成のための Codebook の作成方法について示す。画像の Codebook の作成手順は以下の通りである。1 つの画像について 1 つの Codebook が作成される。

- (1) 各キーワードのグループについて、正解画像 (OK 画像) と不正解画像 (NG 画像) を明確にする。
- (2) 各キーワードのグループの全ての画像から抽出

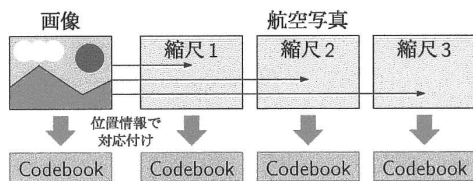


図 4 Codebook の生成

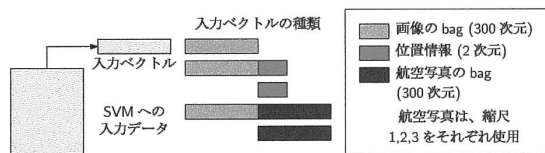


図 5 キーワードごとの SVM への入力ベクトル

された特徴量を用いて、Codebook を作成する。

- (3) 各キーワードから抽出した Codebook を用いて、そのキーワードの画像について、Codebook についてのヒストグラムを作成する。

航空写真の Codebook の作成手順は以下の通りである。

- (1) 各画像における位置情報から、対応する航空写真 (3 種類のズームレベル) を探し、それぞれ $3 \times 3 = 9$ 枚ずつ用意する。
- (2) 各ズームレベル航空写真について、位置 (緯度経度) が中央になるように 256×256 ピクセルの正方形に切り抜く。
- (3) 航空写真について、対応する画像のキーワードに対してズームごとに独立でグループ化し、画像の場合と同様に特徴量、Codebook の順に求める。

図 4 に示すように、この時点で、1 つのキーワードに対して、キーワードの画像、レベル 10、12、14 の各航空写真のグループに関する、4 種類の Codebook を求めることになる。

Codebook を作成は、クラスタリング処理を伴う。1 つのキーワードについて、精度の高い Codebook を作成するには、著しく膨大な量のデータをクラスタリングする必要がある。これは処理時間に対してトレードオフの問題となる。ただし、GRID やランダムで特徴点を抽出した場合、中には重要でない特徴点が含まれていることがある。本実験においては、重要でない特徴点がある程度省くことと、時間を短縮することに配慮し、クラスタリングに用いる特徴点を絞りこむものとする。すなわち、読み込んだ特徴点から、 $\frac{1}{10}$ の確率で特徴点を抽出する。

6.5.2 SVM への入力ベクトルの作成

学習と分類のために用いる SVM への入力ベクトルは、位置情報 (緯度経度) のベクトルと、Codebook に関するヒストグラム (bag) によって構成される。

まず各画像に対して、それぞれのグループの Codebook に関するヒストグラムを作成する。Codebook

画像のみ	画像+位置情報	位置情報のみ
OK データ	OK データ	OK データ
NG データ	NG データ	NG データ
画像+航空写真 縮尺 1	画像+航空写真 縮尺 2	画像+航空写真 縮尺 3
OK データ	OK データ	OK データ
NG データ	NG データ	NG データ
航空写真のみ 縮尺 1	航空写真のみ 縮尺 2	航空写真のみ 縮尺 3
OK データ	OK データ	OK データ
NG データ	NG データ	NG データ

図 6 ベクトルパターンを 2 つのグループに分割

は SIFT 特徴の代表ベクトルを指定されたクラスター数だけ記述したデータであるから、各画像に対応する SIFT 特徴のそれぞれについて、Codebook から「距離が最も近い」ベクトルを探し、そのベクトルに対して投票することによってヒストグラムを作成することができる。ベクトルの距離を計測する尺度として、本研究ではユークリッド距離を用いる。

全ての画像に関してヒストグラムの完成が完了した時点で、それぞれのキーワードにおいて、図 5 に示すように、各ヒストグラム (bag) と位置情報のそれぞれの組み合わせで、各ベクトルを結合したものを作成する。ただし、航空写真については、3 つの各レベルごとにそれぞれ独立して行うので、図 5 では 5 通りの記述のみであるが、実際は 1 つのキーワードに対して 9 通りのベクトルパターンを作成する。

6.5.3 学習と分類

前節において、1 つのキーワードに対して、各画像ごとに、9 通りのベクトルパターンを作成した。各ベクトルパターンのグループのそれぞれについて、図 6 に示すように、対応する画像の手動分類結果に基づき、OK 画像と NG 画像の 2 つのグループに分ける。それぞれのグループから、一定の枚数をランダムに抽出して、Cross Validation の手法によって学習と分類を行う。

7. 実験

7.1 実験データセット

本研究の実験で用いるデータセットは、次の画像からなる。

7.1.1 Flickr 位置情報つき画像

2007 年 6 月から 10 月にかけて、Flickr から収集した約 13 万枚の画像を収集した。各画像には、位置情報 (緯度、経度、18 段階の精度) を含むメタデータがついている。メタデータには、位置情報以外に画像のタイトル、タグ、投稿者 ID、画像の説明の情報を含



図7 実験データセットのOK画像例。上2段は2×2=4枚ずつ、山、海岸、景色、神社、ラーメン、下2段は同様に、ディズニーリゾート、道路、花、東京タワー、鉄道をそれぞれ表す。

表1 分類するキーワードと特徴

キーワード	特徴
景色	遠くが見える景色で、近くに物体がないもの
ラーメン	食べられる状態にあるラーメン
ディズニーリゾート	位置情報などから、ディズニーリゾートで撮影されたと推測できる写真
東京タワー	東京タワーが目立つように写っている写真
山	山頂を含む山の景観が目立つように写っている写真
神社	実際の神社の社殿や、鳥居などの建築と思われるもの
鉄道	電車又は気動車が目立つように写っている写真
道路	道路が全体的に目立つように写っている写真
花	花がアップで撮影されている写真か、写真の殆どが花で占められているような写真
海岸	実際の海岸が写っている写真

む。この画像収集に際しては、位置情報の検索範囲を日本国内に限定した。すなわち、本実験では、日本国内の位置情報を含む画像のみを扱う。

各画像から実験データを作成するために、10種類のキーワード(表1)を考え、各キーワードごとに、メタデータの画像のタイトル、タグ、画像の説明の中に、そのキーワード及びそれに関連する語句を含むような画像を探し、画像の集合を作成する。実験に用いる画像は10キーワードを合わせて10000枚程度である。

キーワードにある画像の集合が、すべてそのキーワードに「適切な」ものであるとは限らない。したがって、それぞれのキーワードの集合に対して、表1の特徴を基準として正解画像を手動で分類する。分類に当たっては、画像とメタデータを参考にした上で総合的に判断する。

7.1.2 航空写真

首都圏から離れた山間部など、位置によっては、高いズームレベルで航空写真が存在しないこともある*1。従って、本実験では、使用する航空写真のズームレベルは図8に示すような3種類とする。Flickrで収集した各写真について、収集した航空写真を用いて、その位置情報を表す航空写真画像を対応づける。

7.2 実験方法

本節では、前述の実験データセットを用いて、学習と分類の実験を行う方法について示す。

7.2.1 SVMへの入力データ

実験データセットから、特徴量を抽出し、Codebookとヒストグラムを作成することによってSVMへの入力データを作成する。これにより、画像と航空写真、位置情報のデータセットから、1つのキーワードについて図6に示すような各グループのデータが作成される。

本実験では、それぞれのグループから、200枚をランダムに抽出して、SVMへの入力データセットを作成する。

7.2.2 Cross Validation

SVMは通常、学習セットで学習を行い、分類セットを分類する。通常ならば学習セットは固定される。本研究ではテストデータ全体をより客観的に評価するため、学習と分類の方法としてCross Validationを用いる。本実験では、5-foldでCross Validationを行う。1つのグループについて、5通りの結果を得る。

*1 これは地域やズームレベルにおいて航空写真の提供元が異なるなどの事情から生じる。

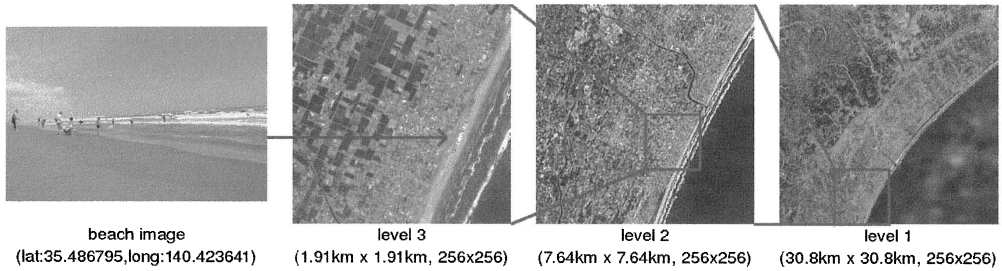


図 8 学習と分類に使用する航空写真

7.3 実験結果及び評価

7.3.1 評価方法

SVM による出力結果から、情報検索の評価と同様に、再現率 (Recall) と適合率 (Precision) を求めることができる。また再現率と適合率の両方を考慮した指標として、 F 値と再現率-適合率グラフを求めることができる。特に再現率-適合率グラフでは両方のトレードオフの関係を検証可能である。

本実験では、学習と分類に Cross Validation を採用している。5-fold で行ったので、上記のような評価方法において、5 通りの結果が出力されることになる。本実験では再現率と適合率の両方を考慮したいところであるが、複数の再現率-適合率グラフを統合し、客観的に検証するのは困難である。そこで、再現率-適合率グラフの面積に相当する平均適合率 (Average Precision) で実験結果を評価する。

7.3.2 平均適合率

SVM による出力値に基づいて、テストデータをソートする。ソートしたデータを最初から順番に読み込み、Positive のデータが出現した時点で、それまでの読み込んだデータの数 m_i と、Positive データの出現頻度 i を記録する。ここで $p_i = \frac{i}{m_i}$ とおく。最後までテストデータを読み込んだときのすべての Positive データの出現数を n とすると、平均適合率 (Average Precision) は、

$$P = \frac{1}{n} \sum_{i=1}^n p_i \quad (2)$$

で計算される。

7.4 実験結果

1 つのキーワードにおいて、9 通りのグループのデータセットをそれぞれ評価する。更にそれぞれのグループを 5 つの fold に分割しているため、5 つの fold のそれぞれの平均適合率を求め、これらの平均を示す。実験結果は、それぞれの平均適合率を 100 倍して有効数字 4 桁で表 2 に示す。表中では、画像を I、位置情報を G、(1),(2),(3) を航空写真のそれぞれのレベルとして、グループの組み合わせを示す。赤太字は、I のみ以外の組み合わせの中で各キーワード毎に最も良かった

結果を示す。diff は、最も良かった結果と、I のみの結果の差を示す。最大で「道路」の場合に、22.15 の平均適合率の向上のみがみられた。一方、「花」の場合は、逆に最も良かった場合でも画像単独の場合 (I) に比べて 3.53 低下してしまった。10 種類の diff の平均値は、11.19 となり、平均的には認識精度が大きく改善された。

8. 考察

位置情報や航空写真の特徴量を、画像の特徴量と統合することで、一部のキーワードを除き、精度の向上が示唆されている。故に本実験の結果より、航空写真と位置情報の特徴量は、画像認識の精度向上を補助する手段である可能性があると見せる。キーワードの中には、位置情報のみでも識別精度が上がるものもあった。

位置情報や航空写真と統合あるいはそれらのみで精度が低下する要因として、OK 画像と NG 画像共に位置情報が一部に集中する場合が考えられる。実際、キーワード「ディズニリゾート」に属する画像の殆んどは、NG 画像でもディズニリゾートの位置情報に集中する傾向が見られた。

8.1 キーワードごとの傾向

「景色」については、「遠くの見える景色」を基準に正解を作成したため、これらの位置情報には、都会や都市部全体的な割合が多い。これらの位置情報に対応する航空写真では、「都市部」の局所特徴量が集中すると考えられる。しかし、都市部の局所特徴量のみでは、風景を識別するのは困難である。したがって風景写真と航空写真の統合した結果が最も精度が高くなると考えられる。

「ラーメン」については、タイトルや説明などのメタデータから、ラーメン店で撮影された画像が多い。「神社」や「東京タワー」と同様に、OK 画像については位置情報が特定の個所に集中しやすい。「ディズニリゾート」も位置情報については同様のことが言えるが、OK 画像と NG 画像の両方で位置情報同様の個所に集中する場合、位置情報によって精度が落ちるようである。

表2 実験結果。本表では画像をI、位置情報をG、(1),(2),(3)を各航空写真の縮尺として略記している。赤字の最高値と画像のみの結果の差をdiffで示す。

キーワード	I	I+G	G	I+(1)	I+(2)	I+(3)	(1)	(2)	(3)	diff
ラーメン	65.49	72.89	73.71	79.10	73.36	67.94	67.78	74.31	56.67	+13.61
神社	56.22	59.04	53.86	56.86	58.15	68.31	62.17	50.41	56.52	+12.09
海岸	74.34	86.15	81.73	73.99	72.73	64.60	74.18	73.16	65.61	+11.81
山	67.29	71.64	69.77	67.64	65.47	67.57	72.38	73.79	70.59	+6.50
景色	78.76	78.22	77.80	79.06	81.92	76.87	65.71	71.62	70.88	+3.16
ディズニー	63.35	57.08	51.27	58.54	53.64	59.05	54.3	62.26	59.55	-1.09
東京タワー	59.62	73.77	65.89	66.84	75.89	66.71	88.17	80.99	85.91	+28.55
鉄道	56.75	61.61	75.47	69.5	60.63	61.48	68.31	62.15	61.31	+18.72
花	75.25	65.65	64.23	63.91	67.1	68.8	65.79	71.72	69.1	-3.53
道路	57.06	64.15	68.48	64.57	79.21	63.68	70.75	66.57	64.06	+22.15

キーワード「山」については、航空写真のみの結果で平均適合率が若干上昇している。一方で、航空写真と山の特徴量を統合したのものについては、画像のみの場合と変わらないか、精度が落ちていると判断できる。この結果から、学習と分類に関して、今回「山」の特徴を決定づけた要因が、山間部の航空写真の特徴量に偏重していると考えられる。

キーワード「海岸」については、位置情報を合わせたデータの結果が最も精度が高い。次いで位置情報のみから学習、分類したデータの精度が高い。一方で、航空写真を用いた場合は精度が向上しなかった。この結果から、「海岸」の特徴を決定づけた要因が、位置情報に偏重していることがいえる。従って、海岸近くで撮影された写真は、被写体が海岸でなくても海岸に分類されることも考えられる。また、航空写真で表現される「海」は、ほぼ均一なため、局所的な特徴を抽出しにくい可能性がある。「鉄道」についても、同様に位置情報のみの特徴で高い精度を得た。

9. 今後の課題

本実験で行った手順が、目標に対する最良の手段であるとは限らない。本実験では限定されたデータセットをもって行ったため、具体的かつ一般的な場合に適用できるよう、データセットを再検討する必要がある。したがってより多くのキーワードについて判定を行うことが課題となる。

また、本実験では特徴量を単純にベクトルの結合で統合した。今後は位置情報と航空写真、そして画像について、特徴量の優良な統合方法を、特徴抽出の段階から再検討する必要がある。

さらに、本実験では比較的低いレベルでの航空写真を利用したため、航空写真と位置情報の範囲を限定し、より高いレベルでの航空写真の利用を含めた実験をすることによって、精度向上の可能性を検討する必要がある。

本研究では、SVMのNG学習データとして、同じキーワードで収集された画像を用いたが、撮影位置について偏りがあったために、適切でなかった可能性が

ある。今後は、多くの位置情報付き画像からランダムで選んだ画像をNG学習データとして利用する実験も行う予定である。

最終的には、位置情報で精度が向上するキーワードとそうでないものを、より明確に判別することが課題となる。

参考文献

- 1) Google Maps. <http://maps.google.co.jp/>.
- 2) Flickr. <http://www.flickr.com/>.
- 3) 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・イメージメディア, Vol.48, No. SIG16 (CVIM19), pp. 1-24, 2007.
- 4) Y.R. Chen, G.D. Fabbriozio, D.Gibbon, S.Jora, B.Renger, and B.Wei. Geotracker: geospatial and temporal RSS navigation. In *Proc. of the 16th International Conference on World Wide Web*, pp. 41-50, 2007.
- 5) S. Ahern, M.Naaman, R. Nair, and J. Yang. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In *Proc. of the International Conference on Digital Libraries*, pp. 1-10, 2007.
- 6) 藤田秀之, 有川正俊, 岡村耕二. 高精度な空間情報付き写真の3次元実空間マッピング. 電子情報通信学会論文誌 A, Vol. J87-A, No.1, pp. 120-131.
- 7) D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol.60, No.2, pp. 91-110, 2004.
- 8) G.Csurka, C.Bray, C.Dance, and L.Fan. Visual categorization with bags of keypoints. In *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 59-74, 2004.
- 9) Thorsten Joachims. SVM^{light}. http://www.cs.cornell.edu/People/tj/svm_light/index.html.