

## Bag-of-Features 表現を用いた エントロピーによる単語の視覚性の分析

川久保 秀敏<sup>†</sup> 柳井 啓司<sup>†</sup>

本研究の目的は、単語概念と画像特徴量の関係性を Web 上の大量の画像データを用いて定量的に分析することである。具体的には、(1) Bag-of-Features 表現を用いた画像領域エントロピーによる単語の視覚性の分析、(2) 位置情報付きの画像の分布を表すジオエントロピーによる単語概念の地理的分布の分析、(3) 画像領域エントロピーとジオエントロピーによる単語の視覚性と地理的分布の関連性の分析、を行った。本研究では、230 語の名詞と、100 語の形容詞について、Web からそれぞれ対応する画像を 500 枚ずつ収集し、これらの分析を行った。分析の結果、“sun” や “rainbow” など空に関する名詞は、他の単語に比べて画像領域エントロピーが小さく、ジオエントロピーが大きい傾向が分かった。一方、地名・地域名や偉人名に関する単語は、ジオエントロピーが小さく、画像領域エントロピーが大きい傾向にあった。

### An Analysis of Word Concepts with Bag-of-feature-based Image Region Entropy and Geo-entropy

HIDETOSHI KAWAKUBO<sup>†</sup> and KEIJI YANAI<sup>†</sup>

The purpose of this study is to analyze a relationship between word concepts and features of images. In this paper, we explore (1) visualness of word concepts with “image region entropy” with the bag-of-features (BoF) representation, (2) geographical distributions of word concepts with “geo-entropy” which is estimated from a large amount of geotagged images, and (3) a relationship between visualness of concepts and geographical distributions. In the experiments, we collected 500 images per concept for 230 noun concepts and 100 adjective concepts from the Web, and applied the proposed BoF-based image region entropy and geo-entropy to all the collected images. From analysis of the results, we found the following tendencies: Word concepts related to “sky” such as “sun” and “rainbow” have smaller image region entropy and larger geo-entropy. In contrast, concepts related to places, areas and historical persons have larger image region entropy and smaller geo-entropy.

#### 1. はじめに

近年、デジタルカメラや WWW (World Wide Web) の普及によって、一般の個人が生活の中でデジタル画像を扱う機会が増加している。デジタル画像を取り扱う際、画像の内容を理解する必要が生じることがある。そのため、計算機が画像の内容を認識し、認識結果を画像への自動処理に生かすことへのニーズが高まっている。現在、計算機による画像内容の自動認識は、限定された状況下で実用化されている。一部のデジタルカメラは、被写体の顔の位置や表情を認識する機能を持っている。この機能は認識対象を人間の顔に限定することで、実用化に足る認識精度を得ている。

一方、制限の無い実世界画像に対して、画像中の物体を認識することを「一般物体認識」<sup>1)</sup> という。一般物体認識の目的の一つとして、画像に対して内容を意味する単語を付与することがある。計算機による一般物体認識では、画像の視覚特徴量やメタデータといった情報をもとに付与する単語を決定する。特徴量と結び付きの強い単語ほど、一般物体認識に用いるのに適している。一般物体認識に適した単語を自動的に選出するためには、単語と画像特徴の関係性を定量化する必要がある。

本研究の目的は、単語概念と画像特徴量の関係性を分析することである。「画像領域エントロピー」<sup>2),3)</sup> を用いて、単語と特徴量の関係性の定量化を行う。SIFT<sup>4)</sup> と Bag-of-Features 表現<sup>5)</sup> を特徴量抽出に用いる。また、単語とジオタグ (画像の緯度経度情報) の関係性についても分析を行う。本研究では 230 語の名詞と 100 語の形容詞を実験対象単語としている。

<sup>†</sup> 電気通信大学 情報工学科

Department of Computer Science, The University of  
Electro-Communications

## 2. 関連研究

単語概念と視覚特徴の関係性を調査した既存研究として、柳井らの研究<sup>2),3)</sup>がある。柳井らは単語概念の視覚性を定量化する手法として画像領域エントロピーを提案し、150語の形容詞について視覚特徴との関係性の強さを調査した。柳井らの研究の結果では、色名の画像領域エントロピーが小さくなる傾向があった。これは特徴量に色特徴を用いたためである。

Koskela らは人手でラベル付けされたデータセットを用いて、概念と特徴量の関係性や、概念間の類似性について調査を行っている<sup>6)</sup>。Koskela らはラベル付き映像のキーフレーム画像から視覚特徴量を算出し、280種類概念(複数の単語で示される概念を含む)との関係性を調査している。Koskela らの研究の結果では、データセット中で出現頻度の高い概念はエントロピーが大きかった。出現頻度が低い概念にはエントロピーが大きいものも小さいものもあった。

柳井らの研究や、Koskela らの研究ではエントロピーに基づく手法で、単語概念と画像特徴量の関係を分析している。本研究もまた、エントロピーを利用した手法で実験と分析を行っている。エントロピーを用いた手法では、単語に関する特徴量のばらつきを分析することができる。一方で、単語概念と特徴量の関係性をモデル化する研究が存在する。Weijer らは色名とピクセル値の関係性をモデル化し、11語の色名について学習実験を行っている<sup>7)</sup>。Ferrari らは縞模様や水玉模様など、部分領域のペアで表現される視覚属性に対応したモデルを提案している<sup>8)</sup>。Gupta らは前置詞や比較形容詞に注目し、2つ部分領域の関係性をモデル化している<sup>9)</sup>。Gupta らは提案したモデルを用いて、部分領域へのラベル付け実験を行っている。

単語概念の視覚性の分析に、位置を取り入れた研究は従来はなく、本研究が初めてである。関連する研究としては、J. Hays らの画像から位置情報を推定する研究のIM2GPSがある<sup>10)</sup>。彼らは、600万枚という大量の位置情報付き画像に対して、類似画像検索をすることによって任意の1枚の画像について、地球上での撮影位置の確率分布を求めることが出来ることを示した。この結果は、画像の位置の分布パターンについて、画像の種類と深く関連があることを示唆している。

## 3. 方針

### 3.1 単語と視覚特徴量の関係性の分析

本研究では、柳井らが提案した「画像領域エントロピー」<sup>2),3)</sup>を用いて単語の視覚性を分析する。画像領域エントロピーを求める手順の概要を以下に示す。

- (1) 単語に関する画像データセットを用意する。
- (2) 画像を領域分割する。

- (3) 部分領域ごとに特徴量を抽出する。
- (4) 単語に関する部分領域(正領域)を自動選出する。
- (5) 正領域に関して、特徴量の分布を算出する。
- (6) 正領域に関する特徴量分布のエントロピーを求め、画像領域エントロピーとする。

柳井らは視覚特徴量として、色・テクスチャ・形状を用いているが、本研究ではSIFT記述子<sup>4)</sup>とBag-of-Features表現<sup>5)</sup>を用いる。また、柳井らは混合ガウス分布を用いて、正領域の選出と特徴量分布の算出を行っている。一方、本研究では正領域の選出にSVM<sup>11)</sup>を利用し、特徴量分布の算出にはpLSA<sup>12)</sup>を用いる。

### 3.2 単語とジオタグの関係性の分析

ジオタグとは写真が撮影された場所の緯度と経度である。Web上の写真共有サービスFlickr<sup>13)</sup>が提供しているWeb API<sup>14)</sup>を用いることで、単語に関連する画像のジオタグを収集することが可能である。本研究では、単語について収集したジオタグのばらつきをエントロピーで評価する。このエントロピーをジオエントロピーと呼ぶことにする。

## 4. 手順

### 4.1 画像領域エントロピーの算出手順

本研究における、画像領域エントロピーを求める手順を図1に示し、各手順についての説明を述べる。

#### 4.1.1 実験画像の収集

本研究では、単語概念に関する画像のデータセットとして、Web上から収集した画像のデータセットを用いる。Web上からの画像収集では、Flickr, YahooのWeb API<sup>14),15)</sup>を用いて画像検索を行う。各単語を検索クエリとし、検索結果の上位のものから順に500枚を収集する。Flickr APIでは検索結果のソート順に

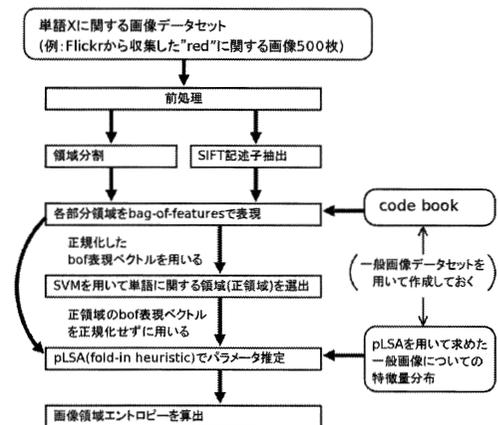


図1 単語 X の画像領域エントロピーを求める手順

relevance を指定し、ジオタグがついていない画像も収集している。Flickr API を用いて収集した画像と、Yahoo 画像検索 API を用いて収集した画像は、別々のデータセットとする。また、人手で作成されたデータセットとして Caltech256<sup>16)</sup> を利用する。

#### 4.1.2 一般画像データセットについて

一般画像データセットは、画像一般を表すためのデータセットである。画像領域エントロピーを求めるときに、結果を比較可能なものとするためには、基準となる潜在トピックモデルが必要である。基準となる潜在トピックを生成するために大量の一般画像データを用いる。Web 上から収集した画像の中から 10000 枚をランダムに選出し、一般画像データセットとした。

#### 4.1.3 前処理・領域分割・SIFT 記述子抽出

画像に対する前処理では、まず長辺が 400 ピクセルを越える画像を、長辺が 400 ピクセルになるように縮小する。そして SIFT 記述子を抽出するためのグレースケール化した画像を作成する。

領域分割は JSEG<sup>17)</sup> を利用して行う。領域の数が多い場合は領域結合処理が行われ、10 個前後の部分領域が得られる。

SIFT++<sup>18)</sup> を用いて、SIFT 記述子抽出をグレースケール化した画像に対して行う。特徴点は SIFT アルゴリズムの DoG 処理によるものではなく、画像中のランダムな座標点 2000 個を特徴点とするようにした。

#### 4.1.4 Bag-of-Features 表現

領域分割で得られた各部分領域を、Bag-of-Features 表現によって特徴ベクトル化する。部分領域に含まれる特徴点から抽出された SIFT 記述子を用いて、code book に関するヒストグラムを作成する。code book は一般画像データセットから抽出した SIFT 記述子を k-means 法でクラスタリングして作成する。code book の要素数は 300 とした。

#### 4.1.5 SVM による正領域の選出

Bag-of-Features 表現ベクトルをもとに、SVM を用いて単語に関係する「正領域」を選出する。

正領域の選出手順を表 1 に示す。単語に関する画像データセットから得られた部分領域が正領域の候補であり、それらの Bag-of-Features 表現ベクトルが正の学習データとして使われる。候補となる領域のうち、1/3 が正領域として選出される。部分領域を単位として選出処理を行うため、正領域が 1 つも選出されない画像も有り得る。選出手順の反復部分は 5 回実行され、5 種類の選出パターンが得られる。

SVM の学習の際には、単語に関係しない負の学習データを与える必要がある。Web 上から収集された画像のうち、500 枚をランダムに選出し、負の学習データを作るための画像データセットとした。これは一般画像データセットの作成方法と同じである。画像一般についてのデータを負とすることで、単語特有の

表 1 正領域の選出手順

step.1 :	正と負の学習データのそれぞれ 1/3 をランダムに選出し、それらを用いて SVM を学習させる。
step.2 :	正と負の全学習データを SVM で評価する。各データの評価結果として、符号付きの値を得る。値が大きいほど正データであると評価されている。
step.3 :	正の学習データのうち、評価結果の値が大きい 1/3 を選出し、正領域とする。選出結果を出力する。
step.4 :	負の学習データのうち、評価結果の値が小さい 1/6 を選出する。さらに、ランダムに 1/6 を選出する。
step.5 :	step.3 で選出したデータと、step.4 で選出したデータを用いて SVM を学習させる。
step.6 :	step.2 に戻る。

Bag-of-Features 表現ベクトルを学習しようと考えた。

#### 4.1.6 pLSA のパラメータ推定

選出された正領域の Bag-of-Features 表現ベクトルを用いて、pLSA のパラメータ推定を行う。正領域の選出パターンは 5 種類得られる。それぞれの選出パターンについてパラメータ推定を行う。パラメータ推定は fold-in heuristics<sup>12)</sup> を用いて行い、一般画像についての  $P(w|z)$  を指定する。 $w$  は Bag-of-Features 表現での code book の要素を示し、 $z$  は pLSA の潜在トピックを示す。

一般画像についての  $P(w|z)$  は、一般画像データセットを用いて求める。一般画像データセットに対して Bag-of-Features 表現ベクトル作成までの処理を行い、pLSA の通常のパラメータ推定で  $P(w|z)$  を求める。潜在トピックの個数は 300 とした。一般画像についての  $P(w|z)$  は、実験全体で共通のものを用いる。

#### 4.1.7 画像領域エントロピー算出

pLSA のパラメータ推定結果をもとに、単語  $X$  の画像領域エントロピー  $H(X)$  算出する。パラメータ推定で求めた確率  $P(z|d)$  を用いて算出する。ここでの  $d$  は正領域を示す。

画像領域エントロピーを求める計算式を式 (1)・式 (2) に示す。式中の  $I$  は正領域の数を示している。5 種類の選出パターンそれぞれについて計算を行い、5 つの画像領域エントロピーを得る。SVM による正領域の選出パターンが 5 種類あるため、5 種類のエントロピーが求まる。その中の最小値を、最終的な画像領域エントロピーとする。

$$H(X) = - \sum P(z_k|X) \log_2 P(z_k|X) \quad (1)$$

$$P(z_k|X) = \frac{1}{I} \sum_i P(z_k|d_i^X) \quad (2)$$

#### 4.2 ジオエントロピーの算出手順

単語に関係するジオタグをヒストグラムにし、エントロピーを求める。ジオタグは Flickr API を用いて収集する。Flickr のジオタグつき画像は、一部のユーザーによって大量に投稿されていることがある。同一

表 2 230 語の名詞

africa, airplanes, alexander, alligator, america, ant, asia, bach, backpack, banana, barbecue, battle, beach, bear, beauty, beaver, bee, beer, beetle, board, boat, bob, book, box, bread, brother, buddha, bug, building, burger, bus, butterfly, cactus, cake, california, canada, candy, canoe, car, castle, cat, cedar, chair, chalk, cherry, chicken, china, circle, city, coffee, coke, color, computer, cookie, coral, crow, dandelion, daughter, desert, desk, dessert, deutschland, dice, dish, doctor, dog, dolphin, dragonfly, dream, duck, eagle, edison, eel, egg, egypt, eiffel tower, election, elephant, elevator, erica, europe, father, fern, field, fireworks, fish, flea, flower, fly, fork, france, frog, fruit, game, gates, giraffe, goat, goose, gorilla, grape, grass, grasshopper, gun, half, ham, hawk, height, helicopter, hibiscus, hornet, horse, house, ice cream, india, insect, italia, ivy, japan, jellyfish, jump, kangaroo, kayak, killer whale, lamp, lavender, lawn, leaf, lemon, level, library, light, lincoln, lion, lizard, locust, love, mangrove, manta, mantis, marriage, mars, milk, mint, monkey, moon, mosquito, moss, moth, mother, mountain, mouse, mozart, museum, mushroom, napoleon, new york, octopus, owl, oyster, palm, paris, park, parrot, party, pen, penguin, people, phone, pine, pizza, plant, pool, pope, potato, pride, pyramid, rabbit, rainbow, rice, rome, rose, salad, salmon, santa claus, school, sea, shakespeare, shark, ship, shrimp, sister, sky, skyscraper, snail, snake, socks, son, sound, spider, sport, square, starfish, steak, sun, sushi, swan, sword, tea, teacher, temple, test, thomas, tiger, toad, tokyo, tool, town, tripod, tulip, tuna, turtle, usa, valley, village, watch, waterfall, whale, wine, worm, zoo
---

表 3 100 語の形容詞

aerial, ancient, antique, bad, beautiful, best, better, big, black, blue, botanical, bottom, bright, brown, cherry, classic, clean, clear, cold, colourful, concrete, cool, crazy, cute, dark, digital, dry, electric, empty, famous, female, first, general, good, grand, gray, great, green, happy, hard, heavy, high, historic, holy, hot, human, iced, interior, international, large, latest, long, male, medieval, military, mobile, modern, more, most, national, natural, nautical, new, nice, old, older, oldest, open, orange, outdoor, pink, present, public, purple, rainy, red, rural, rusted, scenic, second, sexy, short, small, special, sunny, sweet, top, traditional, tropical, twin, underwater, urban, vintage, warm, welcome, white, wide, wild, wooden, yellow
---

表 4 データセットの比較に用いる 30 語の名詞

airplanes, backpack, bear, buddha, butterfly, cactus, cake, canoe, dice, dog, duck, eiffel tower, elephant, fireworks, goose, helicopter, horse, kayak, mars, mushroom, people, pyramid, rainbow, skyscraper, socks, spider, swan, tripod, watch, waterfall
---

表 5 名詞の画像領域エントロピー  $H(X)$

小さい 20 単語 単語	$H(X)$	大きい 20 単語 単語	$H(X)$
sun	3.6497	horse	7.3057
rainbow	4.5538	pizza	7.3071
moon	4.6686	salad	7.3093
dragonfly	4.7550	africa	7.3101
sky	5.1049	japan	7.3387
mantis	5.1897	oyster	7.3435
egg	5.2288	flea	7.3590
airplanes	5.3851	tiger	7.3874
bee	5.4210	rice	7.3906
light	5.4524	rome	7.4013
fly	5.4916	usa	7.4020
coffee	5.6160	backpack	7.4086
bug	5.6407	italia	7.4111
mouse	5.6558	town	7.5177
butterfly	5.6785	santa-claus	7.5431
lemon	5.7096	house	7.5598
dream	5.7173	napoleon	7.5704
lamp	5.7347	school	7.6173
insect	5.7560	lincoln	7.7327
tulip	5.7700	mozart	7.8349

ユーザーによって投稿されたジオタグは、各単語に關して最大 10 個までしか収集しないようにした。

緯度・経度を 10 度ごとに区切って、ヒストグラムのピンとした。よってヒストグラムは 648 次元となる。ヒストグラムは合計が 1.0 になるように正規化される。単語  $X$  についてのヒストグラムからジオエントロピー  $H_{geo}(X)$  を求める式を式 (3) に示す。ヒストグラムの要素を  $b_i$  で表している。

$$H_{geo}(X) = - \sum_i b_i \log_2 b_i \quad (3)$$

ピンの区切る緯度・経度を 5 度ずらすことで、4 通りのヒストグラムが得られる。4 通りのヒストグラムから 4 つのジオエントロピーを求め、その中の最小値を最終的なジオエントロピーとする。

### 5. 実験結果

本研究の実験では 230 語の名詞と、100 語の形容詞を実験対象とした。実験対象単語を表 2 と表 3 に示す。

全実験対象単語について画像領域エントロピーとジオエントロピーを算出した。画像領域エントロピーは Flickr から収集した画像を用いて求めた。

名詞のうち表 4 に示す 30 語については、Yahoo の Web API を利用して収集した画像を用いた場合と、Caltech256 の画像を用いた場合についても画像領域エントロピーを求めた。

名詞と形容詞について求めた画像領域エントロピーとジオエントロピーの一部を表にして示す。まず画像領域エントロピーについて、名詞の中での大小 20 位を表 5 に示し、形容詞の中での大小 10 位を表 6 に示す。次にジオエントロピーについて、名詞の中での大小 20 位を表 7 に示し、形容詞の中での大小 10 位を表 8 に示す。名詞 230 語について、2 種類のエントロ

表 6 形容詞の画像領域エントロピー  $H(X)$

小さい 10 単語 単語	$H(X)$	大きい 10 単語 単語	$H(X)$
orange	5.5608	wooden	7.3469
yellow	5.6780	oldest	7.3728
dark	5.7451	concrete	7.3827
latest	5.8305	older	7.3905
white	5.9033	international	7.3994
clean	5.9623	traditional	7.4268
pink	5.9857	general	7.5015
botanical	6.0937	public	7.5253
happy	6.1351	vintage	7.5972
nautical	6.2085	historic	7.6257

ピーの大小の重なりを図 2 に示す。

表 4 に示した 30 語の名詞について、3 種類の画像データセットを用いて画像領域エントロピーを求めた。その結果を表 9 に示す。各単語について、3 つの画像領域エントロピーの中で最も小さい値を太字にしている。

表 7 名詞のジオエントロピー  $H_{geo}(X)$

小さい 10 単語		大きい 10 単語	
単語	$H_{geo}(X)$	単語	$H_{geo}(X)$
deutschland	0.2602	beetle	5.3225
rome	0.3843	grasshopper	5.3301
tokyo	0.6253	rice	5.3425
paris	0.6730	waterfall	5.3449
eiffel-tower	0.7461	monkey	5.3600
california	0.8776	boat	5.3755
new-york	1.0264	sun	5.4109
italia	1.3105	pool	5.4225
france	1.4833	banana	5.4674
egypt	1.8476	parrot	5.4767
japan	2.2973	sea	5.4936
mozart	2.6904	mother	5.5114
lincoln	2.7962	teacher	5.5417
europa	3.0379	lizard	5.5448
canada	3.2612	fruit	5.5779
castle	3.3948	hibiscus	5.5856
bach	3.4406	ant	5.6147
napoleon	3.4686	coral	5.6565
india	3.4874	fish	5.7831
shakespeare	3.5342	mosquito	5.9759

表 8 形容詞のジオエントロピー  $H_{geo}(X)$

小さい 10 単語		大きい 10 単語	
単語	$H_{geo}(X)$	単語	$H_{geo}(X)$
medieval	3.4364	yellow	5.2554
modern	3.7842	blue	5.2709
new	3.8983	cute	5.2842
rainy	4.1315	small	5.3409
grand	4.1568	beautiful	5.3897
cherry	4.1782	dry	5.4144
big	4.2196	colourful	5.4545
nice	4.2972	tropical	5.5687
public	4.3603	underwater	5.7688
historic	4.4289	traditional	5.8487

画像領域エントロピー		
	小さい 46 単語に含まれる	大きい 46 単語に含まれる
ジオエントロピー	小さい 46 単語に含まれる	tulip france china new-york deutschland africa japan rome usa italia napoleon lincoln morzart pope killer-whale chalk socks shakespeare thomas pride
	大きい 46 単語に含まれる	sun rainbow sky moon airplane bug dream insect beach mosquito beetle beauty banana mangrove ant flower dolphin rice

図 2 エントロピーの大小の重なり (名詞について)

## 6. 考 察

本節では、まず、画像領域エントロピーを算出するために正領域の選出について考察を行う。次に画像領域エントロピーとジオエントロピーの算出結果について考察を行い、最後に単語に関する画像データセットについて考察する。

### 6.1 正領域の選出結果について

SVM を利用して行った正領域選出の結果について

表 9 3 通りの画像データセットから求めた画像領域エントロピーの比較

単語	画像データの収集方法		
	Caltech256	Flickr	Yahoo
airplanes	<b>3.8906</b>	5.3851	6.8635
backpack	<b>5.3903</b>	7.4086	6.7522
bear	7.2956	7.1026	<b>6.7153</b>
buddha	7.6014	<b>6.9686</b>	7.1653
butterfly	6.8613	<b>5.6785</b>	7.1219
cactus	<b>6.3156</b>	6.9601	6.6465
cake	7.6404	<b>7.0314</b>	7.2835
canoe	<b>7.0563</b>	7.2095	7.3451
dice	6.7848	<b>6.3372</b>	6.5667
dog	7.3749	7.1210	<b>7.0530</b>
duck	7.2922	<b>7.1269</b>	7.5398
eiffel-tower	<b>6.1021</b>	6.6226	6.7599
elephant	7.3135	7.1467	<b>6.9260</b>
fireworks	<b>5.7739</b>	6.0937	5.9139
goose	7.3925	<b>6.7513</b>	6.9793
helicopter	<b>6.8465</b>	7.1174	7.6218
horse	<b>6.9009</b>	7.3057	7.0546
kayak	7.2011	<b>6.6147</b>	7.6775
mars	<b>3.1517</b>	6.6767	6.5565
mushroom	<b>6.3381</b>	7.1130	7.0378
people	7.3143	<b>6.8875</b>	7.5739
pyramid	6.8765	<b>6.6736</b>	7.1207
rainbow	<b>3.8686</b>	4.5538	6.1227
skyscraper	<b>5.3588</b>	6.2449	6.8218
socks	<b>6.0510</b>	7.1965	7.5213
spider	<b>5.9760</b>	6.1141	7.2987
swan	7.0208	<b>6.8742</b>	7.8030
tripod	<b>2.9025</b>	6.4607	4.9419
watch	<b>6.5848</b>	6.5987	6.6517
waterfall	6.5121	6.8659	<b>6.3780</b>

考察する。画像領域エントロピーを求めるために選出された正領域の一部を図 3 と図 4 に示す。赤い線が部分領域の境界を示し、正領域には「正」と書かれた緑のラベルが付いている。単語は画像領域エントロピーが小さかった 3 語と大きかった 3 語である。

類似した特徴を持つ領域が数多く出現していれば、それらが正領域として選出されており、正領域選出の処理は概ね狙い通りに動作していると考えられる。

“sun”, “rainbow”, “moon” といった単語では、単語が示す物体や空の領域が多く存在したため、それらが正領域に選出されている。“orange”, “yellow” では、単語が示す色を持つ領域が多かった。しかし今回の実験では、正領域の選出に色特徴を用いていないため、他の色を持つ領域が多く選出されている。“dark” では黒く平坦な領域が多かったため、そうした領域が多く選出されている。

視覚との関係性が弱い単語では、様々な画像が収集される。そのため多様な物体の領域が選出されている。抽象度が高い単語や、多義性を持つ単語では、どういった領域を正領域とするかは難しい問題である。今後の課題として、各領域が正である確率を求めた上で、それを画像領域エントロピーの計算に組み込むことが考えられる。

#### 6.1.1 正領域を選出する反復処理について

本研究の実験では、正領域の選出を反復して 5 回

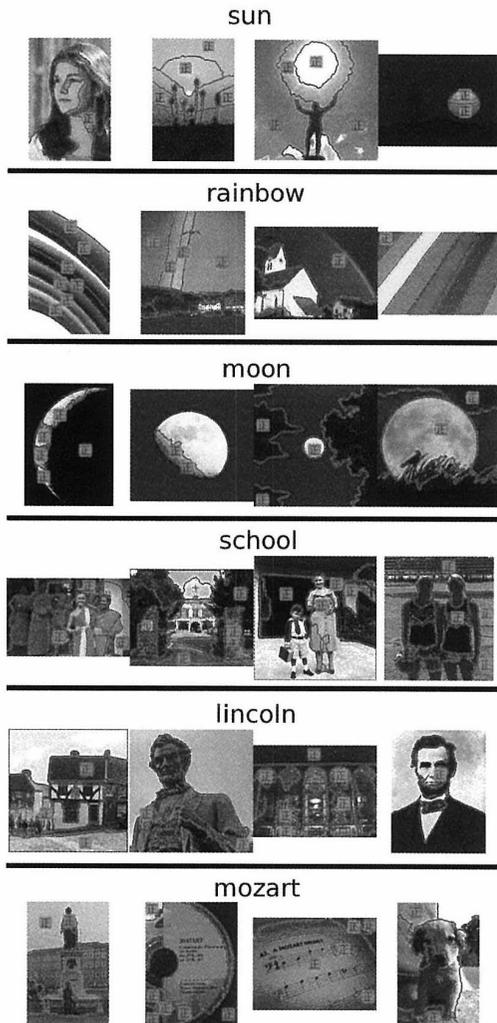


図 3 正領域の選出結果の一部 (名詞)

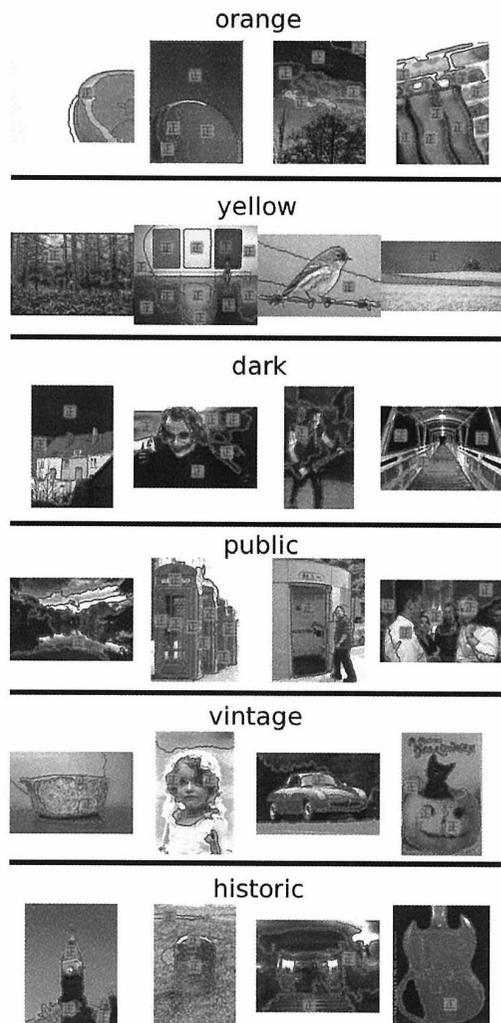


図 4 正領域の選出結果の一部 (形容詞)

行っている。5回の反復のうち、どの回数の選出結果が、最も小さい画像領域エントロピーを与えたのかを調べた(表10)。反復を重ねた場合の選出結果が、最終的な画像領域エントロピーを与える傾向にある。反復処理によって、より小さい画像領域エントロピーを与えるような正領域が得られていることが分かる。

### 6.2 名詞のエントロピーについて

名詞について求めた画像領域エントロピーとジオエントロピーについて考察する。まず、最も特徴的かつ自然な結果として、地名・地域名ジオエントロピーが小さくなる傾向があったことが挙げられる。また、過去の偉人の名前も小さいジオエントロピーを示す傾向にあった。地名・地域名や偉人名といった単語は画

表 10 最終的な画像領域エントロピーを与えた反復回数の統計

反復回数	名詞 230 語		形容詞 100 語	
	Flickr	Yahoo Caltech256	Flickr	
1	4	2 0	2	
2	12	2 1	6	
3	22	1 1	8	
4	57	9 4	26	
5	135	16 24	58	

像領域エントロピーが大きい傾向にあった(図2)。これは地名が示す場所に存在する多様な物体が被写体になりうることや、偉人にちなんだ物体の視覚的多様さを示している。

画像領域エントロピーが小さい単語には、“sun”や“rainbow”といった空に関する単語が挙げられる。こ

これらの単語のジオエントロピーは大きい傾向にあった(図2)。世界中に同じような見目で存在する物は、世界の誰が撮影しても似たような画像となるため画像領域エントロピーが小さくなったと考えられる。反面、世界中に広く存在するためにジオエントロピーが大きくなる傾向にあると考えられる。

“dragonfly”, “mantis”, “bee” といった昆虫の画像は接写撮影がなされているものが多く、背景がぼやけている傾向にあった。ぼやけた領域が多いことが、他の単語に比べて画像領域エントロピーが小さい理由と考えられる。

図2において2種類のエントロピーがともに小さいと分類された“tulip”について考える。“tulip”に関して収集した画像の一部と、ジオタグの分布を図5に示す。“tulip”に関するジオタグは、イギリスやオランダに多く分布している。チューリップの花の色には複数の種類がある。今回の実験では色特徴を使っていないため、小さい画像領域エントロピーが得られたと考えられる。

2種類のエントロピーがともに大きいと分類された“dolphin”, “rice”について考える。収集した画像の一部と、ジオタグの分布を図6と図7に示す。“dolphin”のジオタグは海に面した場所に広く分布している。人工物やマークとしてのイルカの画像も収集されたが、本物のイルカを撮影した写真が多かった。写真の撮影シーンは船上や水中、水族館のショーと多様であり、人間が同時に写っている場合もあった。以上のような理由で画像領域エントロピーが大きくなったと考えられる。

“rice”のジオタグは、日本やタイなどの米を主食とする国を中心に、世界に広く分布している。米料理の見た目が多様であることや、英単語の“rice”が米と稲の両方を示すことが影響し画像領域エントロピーが大きくなったと考えられる。

### 6.3 形容詞のエントロピーについて

形容詞から得られた画像領域エントロピーやジオエントロピーは、名詞から得られたものに比べて、小さい値が少なかった。形容詞は、物体名を示す名詞よりも抽象的であるため、名詞に比べて小さいエントロピーが得られない傾向にあると考えられる。

色を示す単語の画像領域エントロピーが小さい傾向にあった(表6)。ただし色名については、正領域の選出が正しく行っていないと言える(図4の“orange”, “yellow”)正領域を選出する際に色特徴も用いるようすれば、正しく選出できるようになると考えられる。

### 6.4 単語に関する画像データセットについて

表9を見ると、30単語中16単語で Caltech256 の画像を用いた場合の画像領域エントロピーが、他の場合の値より小さかった。caltech256 は人手で作成されたデータセットであり、類似した画像が多く含まれて

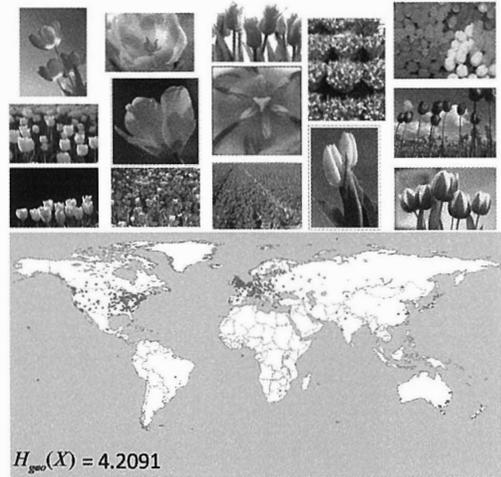


図5 “tulip”に関する画像とジオタグ

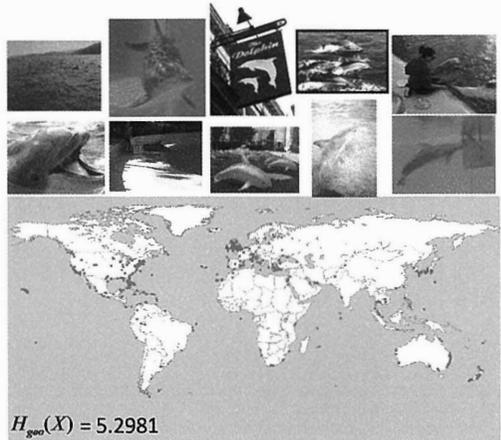


図6 “dolphin”に関する画像とジオタグ

いる上に、単語の多義性が排除されている。そのため画像領域エントロピーが小さくなる傾向が出たと考えられる。

Flickr API を用いた場合と Yahoo 画像検索 API を用いた場合を比較すると、30 単語中 19 単語で Flickr API を用いた場合の方が、画像領域エントロピーが小さい。Yahoo 画像検索 API では幅広い Web ページを検索対象としており、画像が存在する Web ページのテキストを用いた検索が行われる。Flickr API では、Flickr に投稿された画像のみが検索対象であり、Flickr のユーザーが付与したタグや説明文が検索に利用される。こうした画像検索の対象範囲や手法の違いが、画像領域エントロピーの差として現れたと考える。

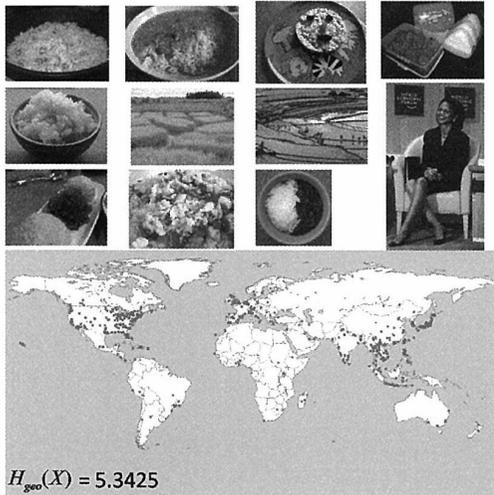


図 7 “rice”に関する画像とジオタグ

## 7. おわりに

本研究では、Web 上から収集した画像を用いて単語の視覚性を分析し、単語とジオタグの関係性についても分析を行った。視覚特徴量として SIFT 記述子による Bag-of-Features 表現を利用し、画像領域エントロピーを用いて分析を行った。単語とジオタグの関係性の分析にはジオエントロピーを用いた。

実験の結果、空に関係する名詞から他の単語に比べて小さい画像領域エントロピーが得られた。また、接写されることが多い虫の画像は、背景のぼやけが共通点となる。それが今回の実験では一種の視覚性として評価された。地名・地域名や偉人名といった単語は、ジオエントロピーが小さく、画像領域エントロピーが大きい傾向にあった。

今回の実験では、正領域を選出するための特徴量と、エントロピーを算出するための特徴量が同一かつ 1 種類であった。色特徴を用いなかったため、色名に関する正領域の選出が正しく行えなかった。数種類の特徴量を利用して正領域の選出を高精度化し、エントロピーを各特徴量ごとに求めれば、単語概念の視覚性についてより詳細な分析を行うことができる。

今回、単語の視覚性と地域性はそれぞれ別々に求めた。今後の課題として、それらを組み合わせた実験を行うことが挙げられる。例えば、画像のジオタグを利用して、世界の各地域ごとに画像を収集し、同一単語の各地域における視覚性を求める実験などが考えられる。この実験を正しく行うためには、実験対象単語に関する画像を、各地域から十分に得る必要がある。

## 参考文献

- 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・イメージメディア， Vol.48, No.SIG16 (CVIM19)，pp.1-24 (2007)。
- 2) 柳井啓司，Barnard, K.：一般物体認識のための単語概念の視覚性の分析，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.SIG10 (CVIM17)，pp.88-97 (2007)。
- 3) Yanai, K. and Barnard, K.: Image Region Entropy: A Measure of “Visualness” of Web Images Associated with One Concept, *Proc. of ACM International Conference Multimedia*, pp.420-423 (2005)。
- 4) Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004)。
- 5) Csurka, G., Bray, C., Dance, C. and Fan, L.: Visual categorization with bags of keypoints, *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp.59-74 (2004)。
- 6) Koskela, M., Smeaton, A. and Laaksonen, J.: Measuring Concept Similarities in Multimedia Ontologies: Analysis and Evaluations, *IEEE Transaction on Multimedia*, Vol.9, No.5, pp.912-922 (2007)。
- 7) Vande Weijer, J., Schmid, C. and Verbeek, J.: Learning Color Names from Real-World Images, *Proc. of IEEE Computer Vision and Pattern Recognition* (2007)。
- 8) Ferrari, V. and Zisserman, A.: Learning Visual Attributes, *Advances in Neural Information Processing Systems* (2007)。
- 9) Gupta, A. and Davis, L.: Beyond Nouns: Exploiting Prepositions and Comparators for Learning Visual Classifiers, *Proc. of European Conference on Computer Vision* (2008)。
- 10) Hays, J. and Efros, A.A.: IM2GPS: Estimating geographic information from a single image, *Proc. of IEEE Computer Vision and Pattern Recognition* (2008)。
- 11) Vapnik, V.: The Nature of Statistical Learning Theory, *Springer-Verlag* (1995)。
- 12) Hofmann, T.: Unsupervised Learning by Probabilistic Latent Semantic Analysis, *Machine Learning*, Vol.43, pp.177-196 (2001)。
- 13) Flickr: <http://www.flickr.com/>。
- 14) Flickr API: <http://www.flickr.com/services/api/>。
- 15) Yahoo 画像検索 API: <http://developer.yahoo.com.jp/search/image/V1/imageSearch.html>。
- 16) Caltech 256 image dataset: [http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/)。
- 17) Deng, Y. and Manjunath, B.S.: Unsupervised Segmentation of Color-Texture Regions in Images and Video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800-810 (2001)。
- 18) Vedaldi, A.: SIFT++, <http://vision.ucla.edu/~vedaldi/code/siftpp/siftpp.html>。