

位置情報付き画像を用いた単語概念の時間変化の分析

ボルドビレグサイハン[†] 及川 雄介[†] 伊藤 祥文[†] 柳井 啓司[†]

[†] 電気通信大学 情報理工学部 総合情報学科 〒182-8585 東京都調布市 調布ヶ丘1丁目5-1

E-mail: †{bileg,oikawa-y,ito-y}@mm.inf.ucc.ac.jp, ††yanai@cs.ucc.ac.jp

あらまし 近年、フォトアルバムサービスやGPS機能付きカメラの普及により、Web上に位置情報付き画像が増加している。特に、Flickrなどのソーシャルネットワークサイトには、テキストタグの付与された位置情報付き画像が対象にアップロードされており、画像に付与されたタグを元に検索することで位置情報付き画像の収集が可能となっている。本研究の目的は、位置情報付き画像を用いて単語概念の位置かつ時間変化との関連性を分析することである。そのために、Web上から大量の位置情報画像を収集しデータセットとして使用する。画像データセットからタグ検索を行い、単語概念に関する画像の集合作る。画像集合を地域かつ時間ごとに部分集合に分割し、各部分集合についてVisualRankアルゴリズムを用いて代表画像を選出する。本研究における地域として“オーストラリア”、“日本”のように国を単位とする。また、撮影時刻を元に、1月から12月までの月単位の時間間隔で時間変化分析をする。代表画像選出には既存の研究と異なってDCNN特徴量を用いることで制度向上が期待できる。地域かつ時間で分割した部分集合ごとの代表画像の位置情報、VisualRank値などを用意して、Webページから読み込むようにして結果を表示する。最後に、地域と時間変化に依存性の高いキーワードと低いキーワードの分類を行う。実験では、単語概念として名詞250語と形容詞100語を用いた。タグ検索で位置情報付き画像を収集し、DCNN特徴量を抽出し地域かつ時間による部分集合に分割し、代表画像を選出した。また、各国内の月別の上位10枚の画像の平均と一年の平均の距離を算出し、距離を元に順位付けを行い、単語概念の時間変化のランキングを行った。

キーワード 代表画像, VisualRank, 地域, 時間, 位置情報

1. はじめに

近年、フォトアルバムサービスやGPS機能付きカメラの普及により、Web上に位置情報付き画像が増加している。特に、Flickrなどのソーシャルネットワークサイトには、テキストタグの付与された位置情報付き画像が対象にアップロードされており、画像に付与されたタグを元に検索することで位置情報付き画像の収集が可能となっている。

本研究では、Web上に存在するこうした大量のタグ付きの位置情報画像を利用して、タグと画像の関係の時間変化について分析を行う。同一のタグに対応する画像は、時間変化によって異なる場合がある。例えば、1月の「木」、4月の「木」、7月の「木」と10月の「木」は、それぞれ視覚的に異なり、「木」に対応する画像特徴量は時間によって異なってくる(図1)。一方、同じ時期でも地域によって季節が異なる場合がある。例えば、地球の南半球の夏と北半球の冬は同じ時期だが季節が異なり、同じ「山」というタグに対応する画像も大きく異なる(図2)。

視覚概念が異なる画像を同じク

そこで、本研究の目的は、Web上にある大量のタグ付きの位置情報画像を用いて、様々なタグについて、このように季節によって画像が変化をするタグと変化しないタグに分類し、変化が大きいタグについてはどのように変化するかを分析することを目的とする。ただし、図2に挙げたように、同じ時期でも地域によって季節や気候が異なるため、単純に世界中の画像を一



図1 月別の「木」の画像例



図2 北半球と南半球の「山」の画像例

緒にして時間変化を分析するのでは、意味のある分析を行うことが難しい。そこで、本研究では、位置情報付き画像を利用することによって、地域毎に特定タグに対応する画像の見目の時間変化を分析することとする。具体的には、地域毎に、各月毎に代表画像を選出し、その代表画像の変化を季節変化として分析することとする。

2. 関連研究

単語概念は地域によって視覚的差があるものがある。視覚に地域差のある単語概念について画像データセットを作成することで一般画像認識の精度向上が期待される。単語概念について地域別の代表画像を選出し、視覚差を定量化することで地域変化に依存性の高い単語概念が分類される。そこで、川久保ら[1], [2]は、代表画像選出するために、画像版のPageRank

アルゴリズムである VisualRank を位置情報画像向けに改良し GeoVisualRank を提案した。実験は名詞 250 語と形容詞 100 語について、世界の 10 都市の座標を用いて位置情報付き画像を収集し、分析を行った。GeoVisualRank の結果を表示するシステムの例を図 3 で表す。

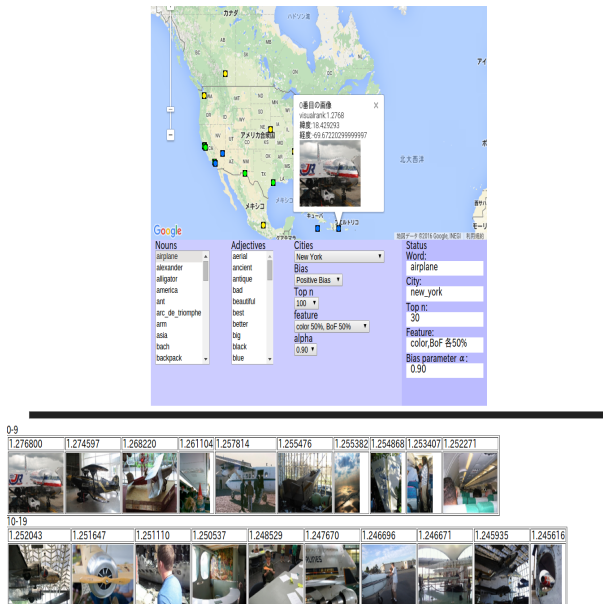


図 3 川久保の研究結果の例 ([1] から引用)。

Jae ら [3] は視覚概念の時間と空間における分析を行っている。歴史的と地域的によって視覚概念がどのような特徴を持つかを調べ、特定の要素を発見した。また、ラベルによりそれらの分類を行っている。

オンラインフォトアルバムサービスではユーザが画像の視覚概念と関連性の高いタグ付けを行う。しかし、時間短縮のために曖昧なタグを付けるユーザや画像の視覚概念と無関係なタグを付けるユーザもいる。そのため不十分なタグやノイズが生じる。Wang ら [4] は画像にメタデータとして付いている位置情報を元に不十分なユーザタグの補完を行った。

3. 手法の概要

本研究の全体的な流れについて説明する。

- (1) 画像収集・集合作成
- (2) 特徴量抽出
- (3) 地域かつ時間で分割・部分集合作成
- (4) 代表画像選出
- (5) 結果表示
- (6) 単語概念の分類

まず、単語概念を元にタグ検索を行い、位置情報付き画像を収集し、画像集合を作成する。画像集合に対して、特徴量を抽出し特徴量データセットを作成する。本研究では画像の特徴量として DCNN 特徴量を使用する。

次に、タグ検索で作成した画像集合を画像の撮影地域と時間で分割し部分集合を作成する。各部分集合に対して、VisualRank アルゴリズムを用いて代表画像を選出する。代表画像を選出す

Layer	1	2	3	4	5	6	7	Output 8
Stage	conv + max	conv + max	conv	conv	conv + max	full	full	full
# channels	96	256	512	1024	1024	3072	4096	1000
Filter size	11x11	5x5	3x3	3x3	3x3	-	-	-
Conv. stride	4x4	1x1	1x1	1x1	1x1	-	-	-
Pooling size	2x2	2x2	-	-	2x2	-	-	-
Pooling stride	2x2	2x2	-	-	2x2	-	-	-
Zero-Padding size	-	-	1x1x1x1	1x1x1x1	1x1x1x1	-	-	-
Spatial input size	231x231	24x24	12x12	12x12	12x12	6x6	1x1	1x1

る際に、VisualRank 値や位置情報、順位などからなるデータセットも作成する。

最後に、用意された VisualRank 値データセットを web ページから読み込み、VisualRank 値の高い順位から画像を位置情報に基づいてマップ上に表示を行う。また、部分集合の上位 10 枚の画像を平均と各上位 10 枚の全体平均の距離を分析し、単語概念の分類を行う。平均同士の距離のばらつきが大きければ時間変化に依存性高い単語概念、ばらつきが低ければ時間変化に依存性の低い単語概念とみなし、分類する。

4. 手法詳細

本研究の手法の詳細について記述する。

4.1 画像収集・集合作成

画像データセットとして Yahoo Flickr Creative Commons 100M(YFCC100M) を用いる。このデータセットは研究用に Yahoo!Labs により公開されている。YFCC100M は約 9930 万枚の画像と約 70 万の動画からなる。YFCC100M データセットには、JPEG 画像の URL、ユーザ ID、撮影時刻、位置情報やタグなどのメタデータが含まれており、そのうち約 566 万枚位置情報付き画像である。本研究で分析対象になっている各単語概念について、YFCC100M からタグ検索を行い、画像を収集し画像集合を作成する。

4.2 特徴量抽出

画像集合に対して DCNN 特徴量を抽出し、特徴量データセットを作成する。本研究では DCNN 特徴量を抽出する際にオープンソースである OverFeat を使用し、1 枚の画像を 4096 次元ベクトルで表現する。

DCNN(Deep Convolutional Neural Network) 特徴量とは、大規模のデータセットで事前に学習したニューラルネットワークの活性化信号を画像特徴ベクトルとして用いたものである。OverFeat は表 1 に示すように、8 層からなるニューラルネットワークで、1000 クラスの約 120 万枚の画像で事前学習を行っている。1 層目から 5 層目までが畳み込み層 (convolution layer) で、残り 3 層が全結合層 (fully-connected layer) になっている。入力画像は 231 × 231 のサイズで、7 層目の活性化信号 (activation signal) の 4096 次元ベクトルを L2 正規化し、特徴量として使用する。

4.3 画像集合の分割

収集した画像によって作成された画像集合を画像の撮影地域と時間によって分割し部分集合を作成する。本研究における地域を“国”単位、時間を“月”単位とし、画像集合を各国かつ月

別で分割する．また，特微量データセットも分割し，特微量部分集合を作成することで，代表画像選出の準備を行う．

4.4 代表画像選出

地域毎の月別代表画像を選出するには PageRank を画像に適用した VisualRank アルゴリズム [6] を利用する．PageRank とは web ページのランク付けを行うためのアルゴリズムであり，web ページ間のリンク構造を行列で表現する．そして反復計算によって各ページのランク値を求める．VisualRank ではリンク構造を行列で表現する代わりに，画像の類似度行列を用いる．VisualRank の基本的な求め方は式 (1) に示す更新式を反復することである．式 (1) では，各画像の VisualRank 値を並べた列ベクトル R を，類似度行列の各列を正規化した行列 S にかけている．VisualRank の前に重要な画像が予想可能な場合は，重要と予想される画像の VisualRank 値の初期値を大きくすることが可能である．

$$R = S \times R \quad (1)$$

式 1 で示した更新式に補正ベクトル P を加えたものが式 2 である．反復ステップごとに，式 1 での結果 P が合成される．PageRank での補正ベクトル P は，リンクを使わずに web ページにアクセスすることをモデル化するためのものである．VisualRank でも更新毎に補正をかけるために P が使用される．補正ベクトル P として一様なベクトルを与えると画像の VisualRank 値が均等化する方向に補正される．更新時に VisualRank 値の合計が変化しないように P の合計と R の合計が等しくなるようにする．

$$R = \alpha(S \times R) + (1 - \alpha)P \quad (0 \leq \alpha \leq 1) \quad (2)$$

4.5 結果表示

VisualRank で求めた代表画像を web ページ^(注1)へ表示する．web ページは，上部のマップと下部の画像の表示部分からなる．下部に月別の代表画像を表示し，各画像の位置情報に基づいてマップ上にマーカーを表示し代表画像とリンクさせる．キーワード選択と地域選択を可能にした動的 web ページを作成する．事前に抽出し用意した VisualRank 値データセットを web ページから読み込み，表示を行う．

4.6 単語概念の分類

各部分集合の VisualRank 値上位 10 枚の画像を用いて分析を行い，キーワードを分類する．各部分集合の類似度が高い場合は，その地域で時間変化に依存性の低い単語となる．一方各部分集合の類似度が低い場合は，その地域で時間変化に依存性の高い単語となる．地域かつ時間変化に依存性の高い単語と低い単語の分類ができる．

各月の平均として，上位 10 枚の画像の特微量ベクトルの平均ベクトルを使用する．式 3 で H_i と表している．また，それらの平均ベクトルの平均を計算し，一年全体の平均として使用する．式 3 で H_y と表している．

$$D_i = d(H_i, H_y) \quad (3)$$

各月の平均ベクトルと一年全体の平均ベクトルのユークリッド距離を算出し，部分集合同士の類似度の評価を行う．そのために，算出したユークリッド距離の集合の分散を用いる．

5. 実験

本研究で行った実験について述べる．まず，実験データセットについて説明する．次に，VisualRank を用いた代表画像選出について述べる．最後に，単語概念に対応する視覚の地域かつ時間に依存性の分析について述べる．

5.1 実験データセット

本研究では，表 2 に示す名詞 250 語と，表 3 に示す形容詞 100 語また表 4 に示す追加名詞 20 語の合計 370 語の単語概念を実験対象とした．名詞 250 語と形容詞 100 語は川久保の研究 [1] を参考に設定した．上記の名詞 250 語と形容詞 100 語の合計 350 語の単語概念について実験を行った結果，収集された画像枚数が少なかったため，追加で自然に関係する名詞 20 語の単語概念についても実験を行った．

単語概念ごとに画像データセット YFCC100M からタグ検索を行い，画像集合を作成した．Flickr では，一部のユーザが類似度の高い画像を大量に投稿していることがある．そのため画像収集時に，同一ユーザが 1 日で投稿した画像を 1 枚に制限した．

収集した画像集合に対し，DCNN 特微量を抽出し，特微量データセットを作成した．特微量抽出の際に overfeat を使用した．

5.2 代表画像の選出結果例

前節で説明した画像集合を画像の撮影地点と時刻を元に分割し，部分集合を作成した．その際に，地域の単位として“国”，時間の単位として“月”を用いた．これによって各単語概念の画像集合を国かつ月別に分割し部分集合を作成し，画像に対する特微量データセットも分割された．

各部分集合である，地域毎の月別画像集合に対し，VisualRank アルゴリズムを使用し，代表画像の選出を行った．その際，式 (4.2) におけるパラメータ α の値は 0.85 に設定した．VisualRank アルゴリズムにより，各部分集合の VisualRank 値を求め，画像に順位付けを行い，VisualRank 値データセットを作成した．

VisualRank について実験結果は，web ページ上に表示した．web ページ下部には，フォームで選択された名詞または形容詞について VisualRank 値データセットを読み込み，またフォームで選択された地域について代表画像上位 10 枚を表示した．web ページ上部の地図には画像の位置情報を元に月別のアイコンを表示した．web ページの上部と下部をリンクさせ，画像をクリックすることでその画像の位置情報を確認でき，アイコンをクリックすることでその位置での画像を確認できた．

“mountain” についての結果: まず，図 4，図 5 に “mountain” をキーワードにした場合の日本とアメリカの実験結果を示す．

図 4 では日本の “mountain” のについて代表画像を月別で表示している．1 月から 3 月まで雪山の画像が現れ，4 月から

(注1): <http://mm.cs.uec.ac.jp/bileg/OFfinal/viewer/>

6月まで上部に雪が積もった山の画像が多く、7月以降の登山シーズンになると山からの風景を表す画像が多い。また、10月以降は雪山が現れる。このことより、冬の山は遠くから撮影した画像で、夏になって登山シーズンになると、山から遠くの風景を撮影した画像である。以上より、キーワード“mountain”は地域“日本”では時間変化によって視覚概念が変動することがわかった。

図5でも同じように冬は雪山、夏は緑の山の画像が多い。以上より、キーワード“mountain”は地域“アメリカ”でも時間変化によって視覚概念が変動することがわかった。

“airplane”についての結果: 次に、図6、図7に“airplane”をキーワードにした場合の日本とアメリカの実験結果を示す。

図6と図7では、時間変化に関係なく同じように飛行機の画像が現れている。人工物である飛行機は時間が変化しても視覚概念の差が少ない。以上より、キーワード“airplane”は地域と時間の変化に依存性の低い単語概念であることがわかった。

“snow”についての結果: 最後に、図8、図9に“snow”をキーワードにした場合の日本とオーストラリアの実験結果を示す。

図8では、日本の“snow”の画像を月別に表示している。図9では、オーストラリアの“snow”の画像を月別に表示している。図8から、地球の北半球の冬である12月、1月と2月に画像枚数が最も多いことがわかる。一方、図9から地球の南半球の冬である7月と8月に画像枚数が最も多いことがわかった。季節によって変化の期待出来る単語“snow”だが、視覚概念の変化が見られず、収集された画像の枚数のみ月別に分布していることがわかった。

6. 単語概念の分析

各月の平均として、上位10枚の画像の特徴量ベクトルの平均ベクトルを使用する。また、それらの平均ベクトルの平均を計算し、一年全体の平均として使用する。各月の平均ベクトルと一年全体の平均ベクトルのユークリッド距離を算出し、部分集合同士の類似度の評価を行う。そのために、算出したユークリッド距離の集合の分散を用いる。

1つの単語概念に対応する位置情報付き画像集合に対して地域と時間で分割を行った。各部分集合の上位10枚の画像の特徴量ベクトルの平均ベクトルを算出し式3の H_i を求めた。ある地域の1月から12月まで各月に10枚以上の画像が存在する時のみ、その地域で分析を行った。各月の平均ベクトルから1年全体の平均ベクトル式3の H_j を求めた。式3によって、各月と全体の平均間のユークリッド距離を求めた。また、ユークリッド距離の分散を求めた。

以上の処理を全ての部分集合に対して行い、分散の大きい順にソートし表5と表6に、分析結果の上位20の単語概念と地域のペアと下位20の単語概念と地域のペアをそれぞれ示した。上位の単語概念の視覚的な変化がその地域で時間変化に依存性高く、下位の単語概念に関して時間変化に依存性が低いことを表している。

表2 250語の名詞

africa, airplane, alexander, alligator, america, ant, arc, arm, asia, bach, backpack, banana, barbecue, battle, beach, bear, beauty, beaver, bee, beer, beetle, big, 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, chicken, china, circle, city, coffee, coke, color, computer, cookie, coral, crow, dandelion, daughter, desert, desk, dessert, deutschland, dice, dish, doctor, dog, dolphin, dragon, dragonfly, dream, duck, eagle, edison, eel, egg, egypt, eiffel, election, elephant, elevator, erica, europe, face, 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, head, height, helicopter, hibiscus, hornet, horse, hospital, house, icecream, india, insect, italia, ivy, japan, jellyfish, jump, kangaroo, kayak, lamp, lavender, lawn, leaf, leg, lemon, level, library, light, lincoln, line, lion, lizard, love, machupicchu, mangrove, manta, mantis, marriage, mars, milk, mint, monkey, moon, mosquito, moss, moth, mother, mountain, mouse, mozart, museum, mushroom, napoleon, new, niagara, octopus, olive, owl, oyster, palm, paris, park, parrot, party, pasta, pen, penguin, people, phone, pine, pizza, plant, playstation, pool, pope, potato, president, pride, pyramid, rabbit, rainbow, rice, rome, rose, salad, salmon, salt, santaclaus, school, sea, shakespeare, shark, ship, shrimp, sister, sky, skyscraper, snail, snake, socks, son, sound, spider, sport, square, starfish, statue, steak, sugar, sun, sushi, swan, sword, tea, teacher, temple, test, thomas, tiger, toad, tokyo, tool, town, train, tripod, tulip, tuna, turtle, uluru, usa, valley, village, watch, waterfall, wave, whale, wii, wine, worm, xbox, 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, open, orange, outdoor, pink, present, public, purple, rainy, red, rich, rural, rusted, scenic, second, sexy, short, small, special, strong, sunny, sweet, top, traditional, tropical, twin, underwater, urban, vintage, warm, welcome, white, wide, wild, wooden, yellow

表4 追加20語の名詞

bamboo, cloud, field, fox, ice, island, lake, rain, river, sakura, seaside, scene, shore, snow, stone, tower, tree, weather, wolf, wood

7. 考 察

本研究の対象となる単語概念 370 語について、位置情報付き画像を収集した。収集した画像を単語概念、地域と時間を元に分割し、部分集合を作成した。各部分集合に VisualRank アルゴリズムを用いて代表画像を選出し、web ページへの表示を行った。

各単語概念の各地域について、全ての月に 10 枚以上の画像が存在する場合のみ分析を行った。画像枚数が不足している場合は分析不可能であるから今回行わなかった。

まず、単語概念 “mountain” に対する実験結果について考察を述べる。1 月から 3 月まで雪山の画像が現れ、4 月から 6 月まで上部に雪が積もった山の画像が多く、7 月以降の登山シーズンになると山からの風景を表す画像が多い。また、10 月以降は雪山が現れる。このことより、冬の山は遠くから撮影した画像で、夏になって登山シーズンになると、山から遠くの風景を撮影した画像である。以上より、単語概念 “mountain” は時間変化によって視覚概念が変動することがわかった。このことより、単語概念 “mountain” のような時間変化に依存して視覚概念も変化する単語概念が存在することがわかった。

次に、単語概念 “airplane” に対する実験結果について考察を述べる。1 年中いつでも同じような「飛行機」の画像が現れている。以上より、単語概念 “airplane” は時間変化によって視覚概念が変動しないことがわかった。このことから、単語概念 “airplane” のような時間変化に依存しない単語概念が存在することがわかった。

最後に、単語概念 “snow” に対する実験結果について考察を述べる。季節的な現象である「雪」について実験を行った結果地域によって各月の画像枚数が変動するが、視覚概念に変化が見られない。このことから、単語概念 “snow” のような時間によって視覚概念を保ったまま、画像枚数が変動する単語概念が存在することがわかった。このような単語概念の時間変化による変動について、本研究で分析不可能であった。

表 5 と表 6 では分析結果の上位 20 の単語概念と地域のペアと下位 20 の単語概念と地域のペアをそれぞれ示した。

表 5 より、時間変化に依存性高い単語概念に、“plant”、“flower”、“tree”、“grass”、などの植物が存在する。また、“field”、“sport”、“rainbow”、などの画像に背景が写っている単語概念が多い。図 10 で時間変化に依存性の高い単語概念の例として “plant” について画像を表示した。以上より、植物や背景を表す単語概念が時間変化によって視覚的に最も変動することがわかった。

表 6 より、時間変化に依存性低い単語概念に、“car”、“train”、“bus” などの乗り物、“building”、“house”、“skyscraper”、“city” などの人工的な建物、“cat”、“dog” などの動物が存在する。図 11 で時間変化に依存性の低い単語概念の例として “train” について画像を表示した。以上より、乗り物や建物のような人工物を表す単語概念が時間変化によって視覚的に変動しないことがわかった。

表 5 分析結果の上位 20

plant	イギリス	0.697906
flower	ドイツ	0.466613
flower	カナダ	0.424134
sport	スペイン	0.389576
tree	イタリア	0.386976
sport	フランス	0.377885
cat	イタリア	0.373871
white	アメリカ合衆国	0.366803
rainbow	アメリカ合衆国	0.359404
field	イギリス	0.355419
grass	イギリス	0.332140
bus	アメリカ合衆国	0.306645
green	イギリス	0.287942
snow	アメリカ合衆国	0.260920
grass	アメリカ合衆国	0.243888
flower	イギリス	0.240290
blue	オーストラリア	0.232362
coffee	アメリカ合衆国	0.229947
fireworks	アメリカ合衆国	0.229654
tree	フランス	0.225323

表 6 分析結果の下位 20

car	ドイツ	0.025102
cat	アメリカ合衆国	0.023655
tree	イギリス	0.023444
city	カナダ	0.023145
people	アメリカ合衆国	0.023051
boat	イギリス	0.022958
car	フランス	0.022756
bus	イギリス	0.021981
dog	アメリカ合衆国	0.021528
city	アメリカ合衆国	0.016707
beach	アメリカ合衆国	0.016371
sea	イタリア	0.016369
house	アメリカ合衆国	0.015683
skyscraper	アメリカ合衆国	0.015535
sky	イギリス	0.012964
train	イギリス	0.010035
building	アメリカ合衆国	0.009543
beach	イギリス	0.008247
car	アメリカ合衆国	0.007954
sky	アメリカ合衆国	0.005550

8. ま と め

本研究では、タグ付きの位置情報画像を利用して、視覚概念の時間変化に関する分析を行った。具体的には、Yahoo Flickr Creative Commons 100M(YFCC100M) データセットからタグ検索で位置情報付き画像を選び、選ばれた位置情報付き画像に対し、OverFeat を使用して DCNN 特徴量を抽出した。次に、画像のメタデータにより、撮影時刻と撮影地点で画像集合を分割し、分割された各部分集合に対し VisualRank アルゴリズムでランキングを行い、VisualRank 値のデータセットを作成した。最後に、用意された VisualRank 値のデータセットを web ページから読み込み、各部分集合の代表画像を表示した。また、各月に 10 枚以上の画像が存在する部分集合に対し、各月の上位 10 枚の画像の平均と 1 年全体の平均との距離を計算し、分散を算出した。各部分集合に対応する単語概念と地域を分散を元にソートし、時間変化に依存性高い単語概念、地域のペアと時間変化に依存性の低い単語概念、地域のペアのリストを作成した。

今後の課題としては、次の 3 点が挙げられる。

まず、本研究の実験対象となった単語概念名詞 250 語と形容詞 100 語について画像収集した結果、画像数が少ない、または、画像が収集されない単語概念が多かった。画像数が少なくて分析不可能となった。このことより、今後課題として、対象となる単語概念の見直しが必要である。

次に、地域を「国」単位で分割した結果、画像数の少ない国が多く現れた。一方、アメリカ、日本、イギリスのように画像枚数の高い国では、「国」単位で画像を分割するには欠点がある。分析対象の地域は画像枚数の大きい地点について半径を定め、画像収集を行うことで、地域の選択ができる。

最後に、単語概念分類の分析では単語概念の時間変化についてのみ行っている。そのため、例えば、“flower” は “ドイツ” と “カナダ” の両国で時間変化に依存性の高い単語である、というような結果が現れる。単語概念の分析に地域を加えることで、単語概念の一般的な時間変化の依存性がわかる。

文 献

- [1] 川久保秀敏, 柳井啓司. 地域別代表画像を用いた単語概念の地域性の分析. 情報処理学会コンピュータビジョン・イメージメディア研究会 (CVIM), 2011.
- [2] H. Kawakubo and K. Yanai. GeoVisualRank: A ranking method of geotagged images considering visual similarity and geo-location proximity. In *Proc. of the ACM International World Wide Web Conference*, 2011.
- [3] Y. Jae, A. A. Hebert, and M. A. Others. Style-aware mid-level representation for discovering visual connections in space and time. In *Proc. of IEEE International Conference on Computer Vision*, pp. 1857–1864, 2013.
- [4] J. Zhang, S. Wang, and Q. Huang. Location-based parallel tag completion for geo-tagged social image retrieval. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015.
- [5] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *Proc. of International Conference on Machine Learning*, 2013.

- [6] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, pp. 1877–1890, 2008.



図 4 日本の“mountain”について代表画像とその位置情報を表示した例



図 6 日本の“airplane”について代表画像とその位置情報を表示した例



図 5 アメリカの“mountain”について代表画像とその位置情報を表示した例



図 7 アメリカの“airplane”について代表画像とその位置情報を表示した例

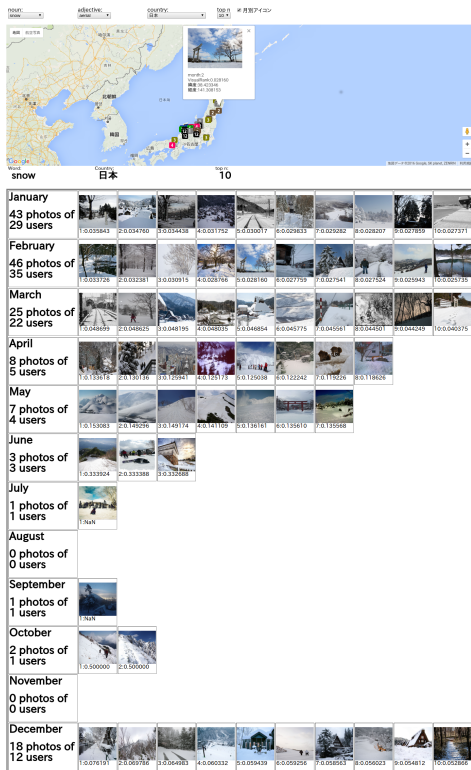


図 8 日本の“snow”について代表画像とその位置情報表示した例

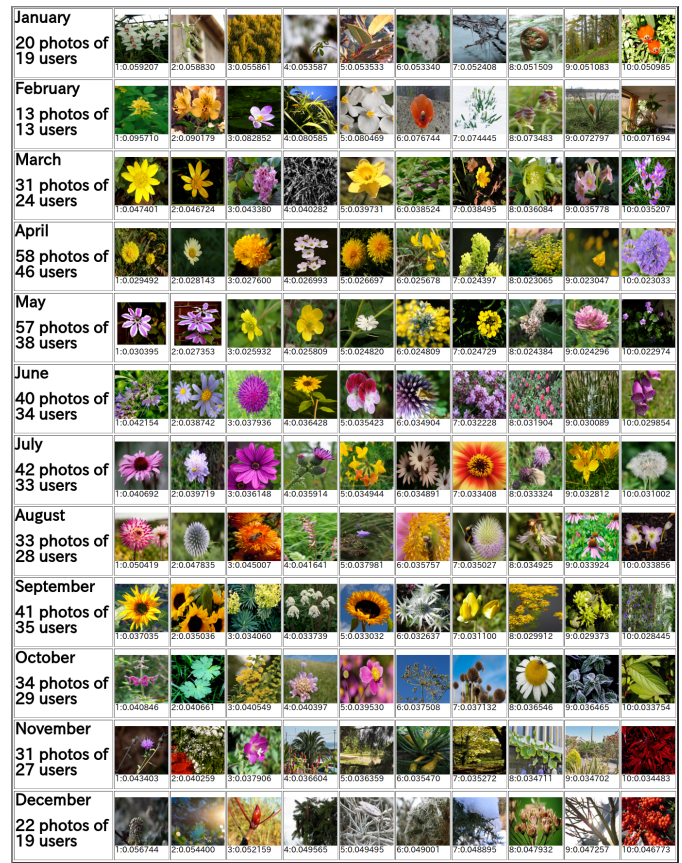


図 10 分析結果の上位の単語概念“plant”について画像を表示した例

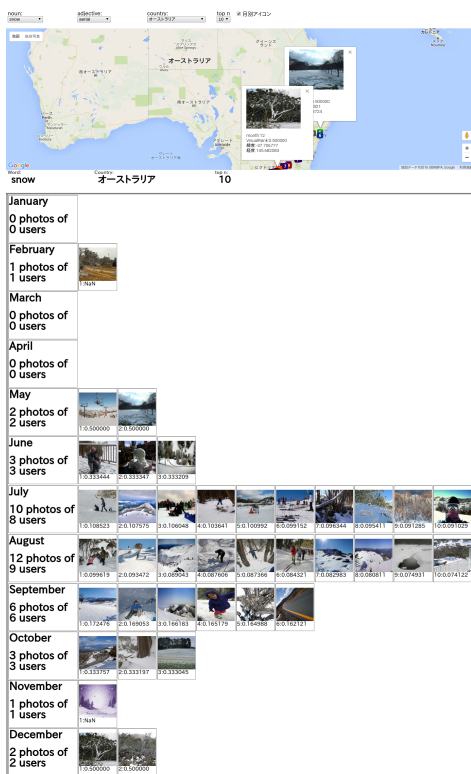


図 9 オーストラリアの“snow”について代表画像とその位置情報表示した例

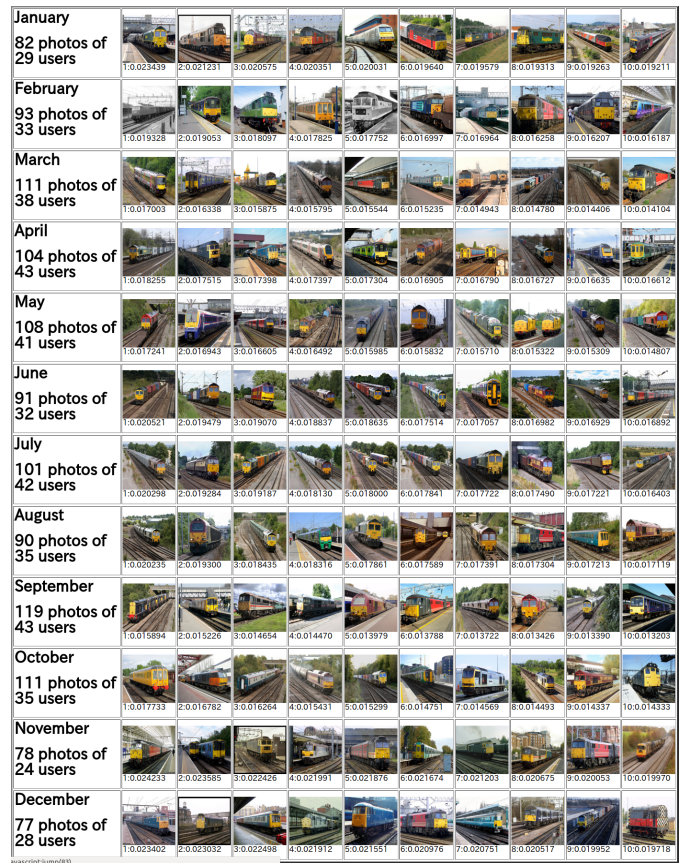


図 11 分析結果の下位の単語概念“train”について画像を表示した例