

地域別代表画像を用いた単語概念の地域性の分析

川久保 秀敏^{†1} 柳井 啓司^{†1}

本論文では、単語概念に対応する視覚の地域性を分析するために、地域別代表画像の選出と分析を行う手法を提案する。提案手法では、まず、単語概念に関する代表位置情報付き画像の集合を用意し、Mean-Shift法によって単語概念に関連する位置座標を複数決定する。そして画像の位置情報と視覚特徴量の両方を考慮する画像ランキング手法“Geo-VisualRank”によって地域別の代表画像を選出する。選出された地域別代表画像の部分領域特徴の分布を分析し、単語概念に対応する視覚の地域差の大きさを求める。そして、地域別代表画像の視覚差の大きさを示す指標値を、pLSAによる潜在トピックモデル化とJSダイバージェンスによる距離計算をもちいた手法で算出する。

実験では、単語概念として250語の名詞と100語の名詞を利用し、各単語概念について視覚の地域性を分析した。その結果、特異な代表画像を持つ地域を含んだ単語概念で大きな指標値が得られた。

An Analysis on Regionality of Word Concepts Using Regional Representative Images

HIDETOSHI KAWAKUBO^{†1} and KEIJI YANAI^{†1}

In this paper, we propose a new method to analyze visual regionality of word concepts using regional representative images. The proposed method consists of the following steps: (1) gathering geotagged images from Flickr regarding given keywords, (2) computing representative geographical coordinates using Mean-Shift clustering, (3) evaluating regional representativeness of each images using GeoVisualRank, (4) calculating the scores of visual regionality using pLSA and JS-divergence. We made experiments about 250 noun concepts and 100 adjective concepts. As results, it turned out that the regionality score of the concepts having unique regional representative images became larger.

1. 研究の背景

1.1 一般画像認識

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。

一方で、多様な単語概念を認識対象とした一般画像認識の高精度化は未だ困難である。計算機による一般画像認識では、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。精度の高い一般物体認識を実現するためには、学習データセットの質が重要である。単語概念についての代表的な特徴量を表現する画像データセットが望ましい。

1.2 位置情報付き画像

近年、Web上のアルバムサービスやGPS機能付きカメラの普及に伴い、撮影地点の緯度経度情報がメタデータとして付与された「位置情報付き画像」がWeb上に増加している。

またFlickrなどの写真共有を行うソーシャルサイトではユーザによる画像へのタグ付けが行われており、タグ検索による位置情報付き画像の収集が可能である。

2. 目的

我々は、画像の位置情報を一般画像認識向けのデータセット構築に役立てるための手法を研究している¹⁾。単語概念に対応する「代表的な視覚特徴」は地域や文化によって、異なる場合がある。例えば、アジアの家とアメリカの家、中東の家、アフリカの家は、それぞれ見た目が異なり、「家」に対応する画像特徴量は地域によって異なってくる(図1)。

視覚に地域差のある「家」などの単語概念では、地域別のデータセットを作成することで、認識精度の向上が期待できる。

地域別の代表画像データセットを構築するためには、単語概念に対応する地域別代表画像を選出するが必要であり、地域別にデータセットを作成することが有効な単語概念を自動抽出するためには、単語概念に対応する視覚の地域差を定量化することが必要である。

本研究では、まず、位置情報付き画像の集合から地域別の代表画像を選出する。そして、地域別代表画像を用いて単語概念に対応する視覚の地域差を定量化する。

^{†1} 電気通信大学大学院 情報工学専攻

Department of Computer Science, The University of Electro-Communications



図 1 各地の家の画像例

3. 関連研究

3.1 画像の撮影位置情報の利用

画像認識分野への位置情報の利用に関連する研究として、Hays らの画像から位置情報を推定する研究の IM2GPS がある²⁾。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。

世界各地のランドマークを認識する研究として、Zheng らの研究や³⁾Crandall らの研究⁴⁾がある。これらの研究では、大量の撮影位置情報をクラスタリングすることで、代表的な都市とランドマークの位置を推定している。そして各地に対応する典型的なランドマークの視覚特徴を求めている。

3.2 VisualRank

画像検索のために、画像集合の中から代表的な視覚特徴をもつ画像を選出する手法が研究されており、画像間の近傍関係が用いられる。

Jing らによって提案された VisualRank⁵⁾ は PageRank アルゴリズム⁶⁾ を画像に適用したものであり、画像間の類似度をもとに画像のランキングを得る手法である。Jing らは、VisualRank をテキストベースの画像検索結果を修正するために用いている⁵⁾。画像の類似度には SIFT 特徴⁷⁾ のマッチ数を使い、商品画像やランドマーク画像を対象にした実験が行っている。

3.3 我々の過去の研究

我々は過去に、単語概念と画像特徴量の関係性を定量化する研究として、画像領域エント

ロピーと位置情報のエントロピー（ジオタグエントロピー）を別々に定量化する実験を行った¹⁾。エントロピーに基づいて単語概念の視覚性と地理的分布を定量化することで、「家」のように地域によって異なる視覚特徴を持つ単語概念と、「空」のような世界中どこでも視覚特徴に大きな変化がない単語概念を、から自動的に発見するための手法を提案した。230 語の名詞と 100 語の形容詞を実験対象とし、Web 上から収集した位置情報付き画像を用いて実験を行った。各単語概念について画像領域エントロピーとジオタグエントロピーをそれぞれ算出し、2 種類のエントロピー値の関係性を調査した。実験の結果、空に関する名詞や、形に多様性の少ない道具の名前から、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一方で、地名・地域名や偉人名といった単語は、ジオタグ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形容詞は、名詞よりも抽象度が高く名詞に比べて小さいエントロピー値が得られなかった

4. 提案手法の概要

本研究の提案手法の概要を図 2 に示す。まずオンラインアルバムサービス Flickr⁸⁾ より、分析対象の単語概念についてタグ・テキスト検索を行い、位置情報付き画像を収集する。

そして、各画像を SIFT 記述子による Bag-of-Features 表現で視覚特徴ベクトル化し、提案手法である GeoVisualRank により撮影位置情報と視覚特徴量を用いて地域別代表画像ランキングを求め、上位画像を選出する。

GeoVisualRank では、地理上の一点の座標を注目点パラメータとして与え、その地点に注目した代表画像の選出が行われる。パラメータとして与えた地理上の一点により近い場所で撮影され、なおかつ代表的な視覚特徴量を持つ画像が上位にランク付けされる。GeoVisualRank を評価するための過去の我々の実験^{9),10)} では、手動で決めた世界の 10 都市の座標をパラメータに用いたが、位置情報付き画像の地理的分布には単語ごとに違いがあり、単語概念ごとに代表的な座標パラメータを自動決定する必要があった。

そのため、収集した画像の位置情報を Mean-Shift 法でクラスタリングし、クラスタ中心の座標をパラメータとして用いる。

選出された地域別代表画像について pLSA による潜在トピックへの帰属確率ベクトルをもとめ、地域ごとに平均し、地域の表現ベクトルとする。地域表現ベクトル間の距離を JS ダイバージェンスによって算出し、地域差の大きさを分析する。

以上の処理によって、地域によって代表画像の視覚特徴に大きな差がある概念と、どの場所でも同じような代表画像が得られる概念に分類することが可能となる。

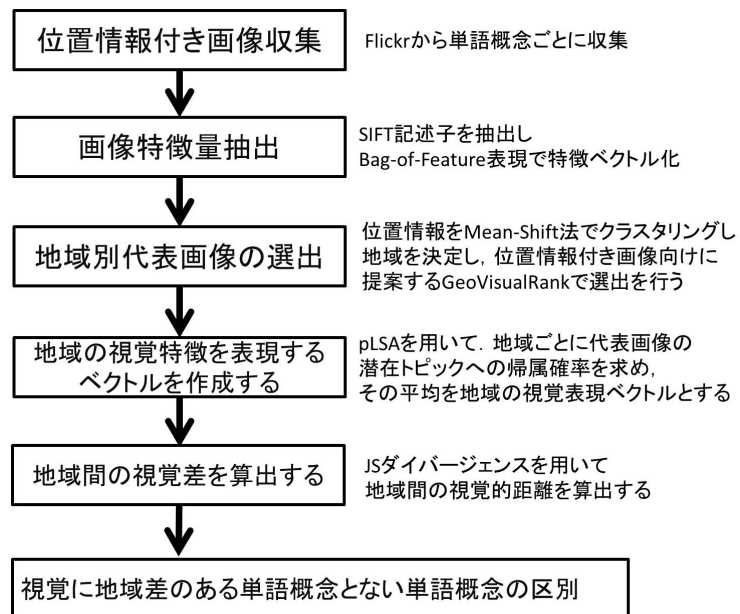


図 2 本研究の手順

5. 提案手法

本節では、提案手法の各手順についての説明を記述する。

5.1 位置情報付き画像の収集

Flickr の提供する WebAPI¹¹⁾ を利用し、Flickr 上の画像を検索・収集する。

本研究では単語概念ごとに Flickr 上の画像を最大 2000 枚ずつ収集し、各単語の位置情報画像データセットとした。Flickr では、一部のユーザーが非常に類似した画像を大量に投稿していることがある。そのため本研究では画像を収集するに当たって、同一ユーザーが投稿した画像の枚数を 20 枚に制限した。

5.2 画像特徴量抽出

本研究では、画像特徴量には、SIFT 記述子⁷⁾ による Bag-of-Features 表現¹²⁾ を利用した。画像を局所特徴量の出現ヒストグラムで表現する Bag-of-Features 表現は、幅広いカテゴリの画像を対象とした場合の表現性の高さから、一般画像認識の分野で頻繁に用いられる。

Bag-of-Features 表現では共通したコードブックを用いることで、異なる画像の表現ベクトルを比較することが可能となる。本研究では収集画像からランダムに選出した 5000 枚の画像を用いて共通したコードブックを作成し、比較可能な Bag-of-Features 表現を求めた。コードブックの要素数は 500 とし、画像は 500 次元の Bag-of-Features 表現ベクトルで表現される。特徴量ベクトルの 1-norm が 1 になるように正規化して利用した。

5.3 代表地域を表現する注目点の決定

本研究での利用する代表画像選出手法である GeoVisualRank では地理上の一点の座標を注目点パラメータとして与え、その地点に注目した代表画像の選出が行われる。

GeoVisualRank を評価するための我々の過去の実験^{9),10)} では、手動で決めた座標をパラメータに用いたが、本研究では、収集した画像の位置情報を Mean-Shift 法でクラスタリングし、クラスタ中心の座標をパラメータとして用いる。

5.4 地域別代表画像の選出

5.4.1 VisualRank

Jing らによって提案された⁵⁾ は PageRank アルゴリズム⁶⁾ を画像に適用したものであり、画像間の類似度をもとに画像のランキングを得る手法である。テキストベースの画像検索によって得られた画像を視覚特徴によって再ランク付けする手法の 1 つである。

各画像の VisualRank 値からなる列ベクトル R は式 (1) を反復し、収束させることで得られる。

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

式 (1) 中の S は、画像類似度行列の各列を正規化したものである。各列の合計が 1 になるように正規化することで、VisualRank 値の合計が変化しないようにする。

P は補正ベクトルであり、PageRank アルゴリズムではリンク以外での Web ページへのアクセスをモデル化している。Jing らは P に均一なベクトルを用いる手法と、画像収集時のテキストベース画像検索での順位に基づくバイアスペクトルを用いる手法を提案した。

補正の強度はパラメータ α によって調整する。多くの場合、 $\alpha \geq 0.8$ を満たす値が設定される。

5.4.2 GeoVisualRank¹³⁾

我々は、VisualRank を適用する画像を位置情報付き画像に限定し、位置情報に基づくバイアスペクトルを生成し、利用する GeoVisualRank を提案した^{9),10),13)}。GeoVisualRank ではランキングを生成する際に、地理上の一点を「注目点」として指定する。注目点に近い座標で撮影された画像のランキング値が大きくなるようにバイアスを作成することで、代表

的な画像特徴を持ち、撮影された地点が注目点に近い画像ほど上位にランキングされる。

Jing らは画像の類似度を求めるために、局所特徴量である SIFT 特徴のマッチ数を用いている。これは商品やランドマークといった特定の見た目を持つ物体の画像について有効である。本研究ではより多様な見た目を持つ単語にも対応するため、SIFT 記述子による Bag-of-Features 表現を用いている。

画像特徴量ベクトルについての類似度計算には、正規化したヒストグラムについてのヒストグラムインタセクションを用いた。

5.4.3 位置情報に基づくバイアスベクトル

GeoVisualRank では地理上の一点の座標をパラメータとして与え、注目点とする。注目点に近い地点で撮影された画像のランキング値をより大きくするバイアスベクトルを作成し、利用する。バイアスベクトルによる補正は各反復ステップでかかるため、『注目点付近の画像によく類似した画像』の VisualRank 値も間接的に大きくなる。

位置情報によるバイアスベクトル P_i^{geo} は、式 (2) で作成する。本研究では、式 (2) の定数 c は 1 とした。式 (2) の D_i は画像 i の位置情報と注目点の距離である。地球を半径 1 の完全な球体とした場合の球面上の距離を球面三角法で計算する (式 (3))。 $lat_i, long_i$ は画像 i の位置情報であり、 $lat_C, long_C$ は注目点の緯度・経度である。

バイアスベクトルは、VisualRank ベクトル R と同じ 1-norm になるよう正規化して使用する (式 (4))。要素の合計値を R と同じにすることで、反復計算時に VisualRank 値の合計が変化しないようにする。

$$p_i^{geo} = \exp(-cD_i) \quad (2)$$

$$D_i = \cos^{-1}(\sin(lat_i) \sin(lat_C) + \cos(lat_i) \cos(lat_C) \cos(long_i - long_C)) \quad (3)$$

$$P_i^{geo} = \frac{P_i^{geo}}{\|p^{geo}\|_1} \|R\|_1 \quad (4)$$

5.5 単語概念の地域性の分析

地域別の代表画像を用いて、単語概念に対応する視覚的地域差を分析する手法について説明する。まず分析手法の概要を述べ、各手順について説明する。提案手法では、GeoVisulrank を用いて地域別に求めた代表画像ランキングの上位画像を地域別代表画像として扱う。

視覚的地域差を分析するために、まず、pLSA (probabilistic latent semantic analysis)¹⁴⁾ を用いて、代表画像の特徴量ベクトルから潜在トピックへの帰属ベクトルを求める。

そして、地域内の代表画像について、潜在トピック帰属ベクトルを平均することで地域の潜在トピック帰属ベクトルを求める。地域間の視覚的距離は、地域の潜在トピック帰属ベクトルの JS ダイバージェンスで算出する。

単語概念内に、他の地域との視覚的距離が大きい地域が存在すれば、その単語概念に視覚的地域性があると期待できる。そこで、単語概念内の各地域について、他地域との視覚的距離の平均をもとめ、その最大値を単語概念の視覚的地域性を表す指標とした。

5.5.1 pLSA によるトピック帰属表現

pLSA (probabilistic latent semantic analysis) は、Hofmann によって提案された生成モデルであり、元々は統計的言語処理のためのモデルである¹⁴⁾。以下の説明では、文章を $d_i (i = 1, 2, \dots, I)$ で表す。同様に単語を $w_j (j = 1, 2, \dots, J)$ で、潜在トピックを $z_k (k = 1, 2, \dots, K)$ で表す。文書 d_i における単語 w_j の発生確率は、潜在トピック z_k を用いて式 (5) のように表される。

$$P(w_j|d_i) = \sum_{k=1}^K P(w_j|z_k)P(z_k|d_i) \quad (5)$$

画像認識分野では、文書を画像に、単語を特徴量に置き換えて利用される。Bag-of-Features 表現ベクトルに対して利用する場合、単語 w_j はコードブックの要素である。本研究では潜在トピック数 K は 300 として利用した。

Fold-in heuristics¹⁴⁾ は pLSA のパラメータ推定の際に、 $P(w_j|z_k)$ を指定の値で固定する手法である。

複数のデータセットで各々にパラメータ推定を行う場合、fold-in heuristics によって共通の $P(w_j|z_k)$ を指定すれば、パラメータの推定結果が比較可能となる。

$P(w_j|z_k)$ は潜在トピックの内容を示していると言えるので、fold-in heuristics で共通の $P(w_j|z_k)$ を指定することは、共通の潜在トピックを用いたパラメータ推定を行うことであると言える。共通の潜在トピックを用いることで推定結果の $P(z_k|d_i)$ が比較可能となる。

本研究では収集画像からランダムに選出した 5000 枚の画像を用いてあらかじめ共通した潜在トピックを作成し、fold-in heuristics を適用することで比較可能な $P(z_k|d_i)$ を求めた。

そして、地域 R の代表画像ランキング上位 N 枚の潜在トピック帰属ベクトルを平均することで、地域 R の潜在トピック帰属ベクトル $P(z_k|R)$ が算出される (式 (6))。式 (6) の

d_i^{TOP} は地域 R の代表画像ランキングで上位 i 番目の画像を示す。

$$P(z_k|R) = \frac{1}{N} \sum_{i=1}^N P(z_k | d_i^{TOP}) \quad (6)$$

5.5.2 JS ダイバージェンスによる距離計算

JS ダイバージェンス (Jensen-Shannon divergence) は 2 つの確率分布間の距離尺度である。JS ダイバージェンスは、対称性がないために距離尺度としては不完全な KL ダイバージェンス (Kullback-Leibler divergence) に対称性を与えた尺度となっている。2 つの確率分布間の類似性が高いほど値は小さくなる。2 つの確率分布 P, Q に対してそれぞれの算出方法は次式ようになる。

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (7)$$

$$D_{JS}(P||Q) = \frac{D_{KL}(P||(P/2+Q/2))}{2} + \frac{D_{KL}(Q||(Q/2+P/2))}{2} \quad (8)$$

地域間の視覚的距離は、地域の潜在トピック帰属ベクトルの JS ダイバージェンスで算出する。地域 R_1, R_2 間の視覚的距離 $D_R(R_1||R_2)$ の算出を式 (9) 示す。

$$D_R(R_1||R_2) = D_{JS}(P(z_k|R_1) || P(z_k|R_2)) \quad (9)$$

5.5.3 単語概念に対応する視覚的地域性の推定

単語概念内に、他の地域との視覚的距離が大きい地域が存在すれば、その単語概念に視覚的地域性があると期待できる。

単語概念内の各地域について、他地域との視覚的距離の平均をもとめ、その最大値を単語概念の視覚的地域性を表す指標値とした。この指標値が大きいほど、特徴的な視覚を持つ代表地域が存在する単語概念と期待できる。

単語概念 C の視覚的地域性の指標値 $VisualRegionality(C)$ の算出方法を式 (10) に示す。式 (10) の R_C は単語概念 C についての代表地域の集合を示す。

$$VisualRegionality(C) = \max_{R_i \in R_C} \left\{ \frac{1}{|R_C| - 1} \sum_{j=1}^{|R_C|} D_R(R_i || R_j) \right\} \quad (10)$$

6. 実験

本章では、本研究で行った実験について記述する。まず実験で用いたデータセットについて説明し、それから単語概念に対応する視覚的地域性の分析実験について述べる。

6.1 実験データセット

本研究では、名詞 250 語と形容詞 100 語の合計 350 語を実験対象の単語概念とした。名詞は、身の回りの物体の名称、生物名、固有名詞などを人手で設定した。形容詞は、単語概念と視覚特徴の関係性を調査した、柳井らの研究¹⁵⁾を参考にして設定した。

単語概念ごとに Flickr 上の画像を最大 2000 枚ずつ収集し、各単語の位置情報画像データセットとした。Flickr では、一部のユーザーが非常に類似した画像を大量に投稿していることがある。そのため画像を収集するに当たって、同一ユーザーが投稿した画像の枚数を 20 枚に制限した。

6.2 視覚的地域性の分析実験

本節では、単語概念に対応する視覚的地域性の分析実験について述べる。各単語概念について、提案手法によって視覚的地域性を示す指標値を算出した。

また、本実験では表 1 に示すパラメータ値の全組み合わせで実験を行った。

表 1 パラメータとして用いた値のリスト

パラメータ	用いた値のリスト
Mean-Shift 法での半径パラメータ (km)	50, 100, 250, 500, 750, 1000
GeoVisualRank でのバイアスパラメータ α	0.7, 0.75, 0.8, 0.85, 0.9, 0.95
ランキング上位何枚を代表画像とするか	5, 10, 20, 50

Web 上から視覚的地域性の指標値ランキング^{*1}が閲覧可能である。フォームによって選ばれたパラメータに応じたランキング結果が表示される。単語概念をクリックすると使用された地域別代表画像が表示される。

また、選出された地域別代表画像の撮影位置も Web 上^{*2}で閲覧することができる。ページ左部の地図には撮影位置が地域別に色分けされたアイコンで表示される。ページ右部に表示される画像をクリックすると、その画像の位置情報を地図上で確認できる。

*1 <http://mm.cs.uec.ac.jp/kawaku-h/regionalranking/>

*2 <http://mm.cs.uec.ac.jp/kawaku-h/visualregionality/>

6.2.1 結果例 視覚地域性の指標値ランキング

視覚地域性の指標値の算出結果の例として、表 2 に示すパラメータを用いた場合について結果を示す。実験対象とした 350 語のうち、視覚地域性の指標値の大きい 10 語と小さい 10 語を表 3 に示す。

表 2 パラメータ設定

パラメータ	用いた値
MeanShift 法での半径パラメータ (km)	500
GeoVisualRank でのバイアスパラメータ α	0.85
ランキング上位何枚を代表画像とするか	5

表 3 視覚地域性の指標値の大きい 10 語と小さい 10 語

指標値の大きい 10 語		指標値の小さい 10 語	
指標値	単語	指標値	単語
0.00940416	building	0.00605590	california
0.00935062	holy	0.00613669	deutschland
0.00924220	airplane	0.00633069	machu picchu
0.00918682	beach	0.00675465	niagara
0.00912758	half	0.00706835	uluru
0.00911686	field	0.00735098	arc de triomphe
0.00908912	monkey	0.00749106	concrete
0.00908885	whale	0.00750860	chair
0.00908138	eel	0.00752026	book
0.00907977	dark	0.00752867	election

地名とランドマークの指標値が小さいと分かる。地名やランドマークでは、位置情報が狭い範囲に集中するため、世界の地域別のデータセットを作りようがないことを示している。

一方で、大きい指標値が得られた 'building' や 'holy' の代表画像 (図 3, 図 4) をみてみると、図 3 の cluster.4 や cluster.7 のような代表画像に同一物体の画像を複数含むクラスタが大きい指標値を与える要因となっている。

特に、すべての代表画像が同一物体の画像で構成されている cluster.7 のような場合は、ノイズと判断する必要がある。特定地域についての特有な外観をもつ有用な画像を残しつつ、ノイズとなるような画像を除去する手法が必要である。

今回、Flickr 画像を収集する際に、同一ユーザから投稿された画像の枚数の制限を 20 枚とした。今後は、同一ユーザの画像の取得数をさらに制限する必要があると考えられる。



図 3 'building' の地域別代表画像

6.2.2 半径パラメータによる代表地域画像の変化

代表画像の撮影位置を確認すると、Mean-Shift の半径パラメータが大きいほど、世界の幅広い地域が代表地域として選ばれる傾向があった (図 5, 図 6, 図 7)。本研究では、所属クラスタ数の多い上位 10 クラスタを代表地域クラスタとした。そのため半径パラメータが小さい場合、撮影位置データの密度の高い地域が複数に分割され上位クラスタを占めると考えられる。一方で、半径パラメータを 1000 とした場合には、生成されたクラスタの数が 9 になり、代表地域の数 が 10 未満になっている。幅広い範囲から十分な数のクラスタを得るためには、半径パラメータを調節する必要があると考えられる。



図 4 'holy' の地域別代表画像



図 5 'house' の結果例, 半径パラメータ 50

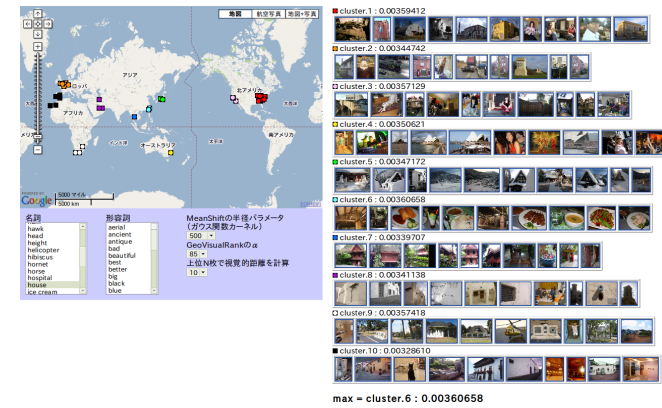


図 6 'house' の結果例, 半径パラメータ 500

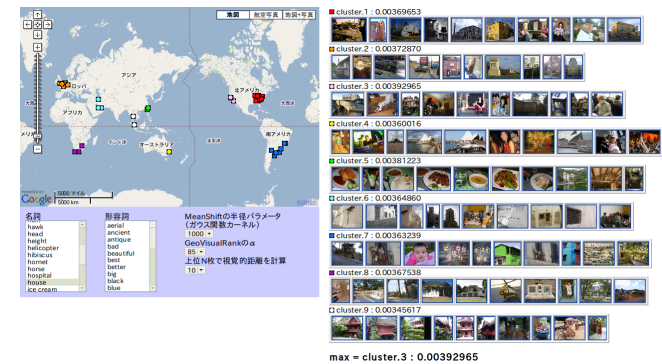


図 7 'house' の結果例, 半径パラメータ 1000

7. ま と め

本研究では、位置情報付き画像向けの代表画像選出手法である GeoVisualRank を提案した。そして、単語概念についての位置情報付き画像データセットから、単語概念の視覚の地域性を定量化する手法を提案した。

350 語の単語概念を実験対象として、Flickr から位置情報付き画像を収集してデータセットを構築した。そして、単語概念の視覚の地域差を定量化する実験を行った。

その結果、単語概念の視覚の地域差を定量化するために提案した手法は、地域ごとに選ばれた代表画像の視覚差を定量化することが出来た。しかし、ほぼ同一の複数の位置情報付き画像がデータセットに存在した場合に、影響を受けてしまうということが分かった。

8. 今後の課題

単語概念の視覚の地域差を定量化するために提案した手法では、今後、重複した画像の影響を受けにくいものに改良していく必要がある。提案手法に与える前の時点でノイズとなるような重複画像を省く方法が考えられる。Flickr 画像については、投稿者 ID や位置情報や撮影日時を調べることによって、過度に類似した画像を入力とすることを防止できると期待できる。

入力データセットのなかに重複ノイズ画像があっても、影響を受けにくいように提案手法を改良できれば、より望ましいと考える。

参 考 文 献

- 1) 川久保秀敏, 柳井啓司. 単語概念の視覚性と地理的分布の関係性の分析. 電子情報通信学会論文誌, Vol.93, No.8, pp. 1417–1428, 2010.
- 2) J.Hays and A.A. Efros. IM2GPS: Estimating geographic information from a single image. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2008.
- 3) Zheng, Y.T. and Zhao, M. and Song, Y. and Adam, H. and Buddemeier, U. and Bissacco, A. and Brucher, F. and Chua, T.S. and Neven, H. Tour the world: building a web-scale landmark recognition engine. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2009.
- 4) Crandall, D.J. and Backstrom, L. and Huttenlocher, D. and Kleinberg, J. Mapping the world's photos. In *Proc. of ACM WWW Conf.*, 2009.
- 5) Y.Jing and S.Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.30, No.11, pp. 1870–1890, 2008,.

- 6) S.Brin and L.Page. The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems*, Vol.30, No. 1-7, pp. 107–117, 1998.
- 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) Flickr. <http://www.flickr.com/>.
- 9) 川久保秀敏, 柳井啓司. Visualrank における位置情報活用の検討. 電子情報通信学会研究会報告: パターン認識・メディア理解研究会, 第 109 巻, pp. 51–56, 2009.
- 10) 川久保秀敏, 柳井啓司. 位置情報を考慮した visualrank による地域別代表画像の選出. 画像の認識・理解シンポジウム (MIRU 2010), 2010.
- 11) Flickr API. <http://www.flickr.com/services/api/>.
- 12) 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.
- 13) H.Kawakubo and Y.Keiji. Geovisualrank: A ranking method of geotagged images considering visual similarity and geo-location proximity. In *Proc. of ACM World Wide Web Conference*, 2011.
- 14) T.Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, Vol.43, pp. 177–196, 2001.
- 15) 柳井啓司, Kobus Barnard. 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.