IS1-23

GeoVisualRankを用いた単語概念の地域性の分析

電気通信大学大学院　情報理工学研究科 総合情報学専攻　川久保秀敏　　樋爪和也　柳井啓司

背景と目的

精度の高い一般物体認識のためには学習データの質が重要

位置情報付き画像がWeb上で増加

◆大量の位置情報付き画像から単語概念に対応する「視覚の地域性」を定量化し、発見する

- ➢地域ごとの「代表的な視覚特徴」を自動抽出
- ➢地域別のデータセットを作成して、認識精度向上へ

手法概要

位置情報付き画像収集 — Flickrから単語概念ごとに収集

↓

画像特徴量抽出 — SIFT記述子を抽出し Bag-of-Feature表現で特徴ベクトル化

↓

地域別代表画像の選出 — 位置情報をMean-Shift法でクラスタリングし地域を決定し，位置情報付き画像向けに提案するGeoVisualRankで選出を行う

↓

地域の視覚特徴を表現するベクトルを作成する — pLSAを用いて．地域ごとに代表画像の潜在トピックへの帰属確率を求め，その平均を地域の視覚表現ベクトルとする

↓

地域間の視覚差を算出する — JSダイバージェンスを用いて地域間の視覚的距離を算出する

↓

視覚に地域差のある単語概念とない単語概念の区別

1. 画像収集

- 単語概念ごとに最大2000枚
- 同一ユーザからは最大20枚に制限

2. 画像特徴抽出

- SIFT
- Bag-of-Features表現
  - ➢500次元

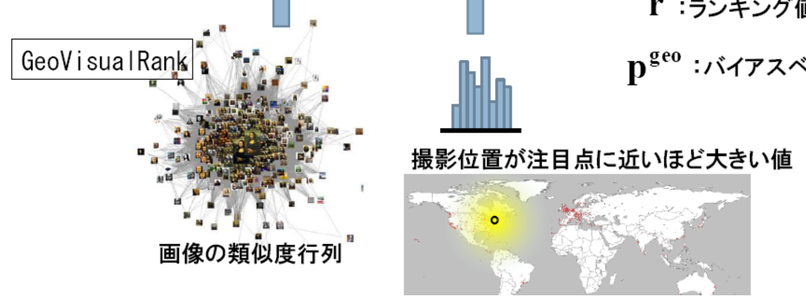
3. 地域別画像の選出

- クラスタリングによる位置情報の決定
- GeoVisualRankによる評価
- 位置情報のバイアスベクトルによる補正
  - ➢クラスタ中心に近いほどランキング値をより大きく

$$\mathbf{r} \Leftarrow \alpha \mathbf{S}^{*} \mathbf{r} + (1-\alpha)\mathbf{p}^{geo} \qquad (0 \le \alpha \le 1)$$

$\mathbf{r}$ ：ランキング値ベクトル

$\mathbf{p}^{geo}$ ：バイアスベクトル



4. 表現ベクトル作成

- pLSAにより潜在トピックへの帰属ベクトルへ変換

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

R　：地域
z_k　：潜在トピック
d_k^{TOP}　：ランキング上位*i*番目の画像

5. 距離計算

- JSダイバージェンス
  - ➢閾値以下なら関係性あり

$$D_{KL}(P \parallel Q) = \sum_i P(i)\log\frac{P(i)}{Q(i)}$$

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

6. 視覚の地域性の推定

- 単語概念内で他地域との視覚的距離を計算
  - ➢単語概念Cの地域性の指標値VisualRegionalityを計算

$$VR(C) = \max_{R_i \in R_C}\{\frac{1}{|R_C|-1}\sum_{j=1}^{|R_C|} D_R(R_i \parallel R_j)\}$$

実験・結果

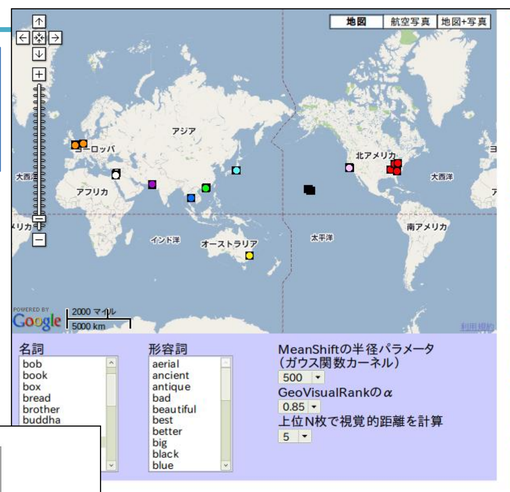
□ 名詞250語、形容詞100語

実行結果　公開用URL
http://mm.cs.uec.ac.jp/kawakuh/visualregionality/

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

半径500、α=0.85、代表画像上位5枚の場合

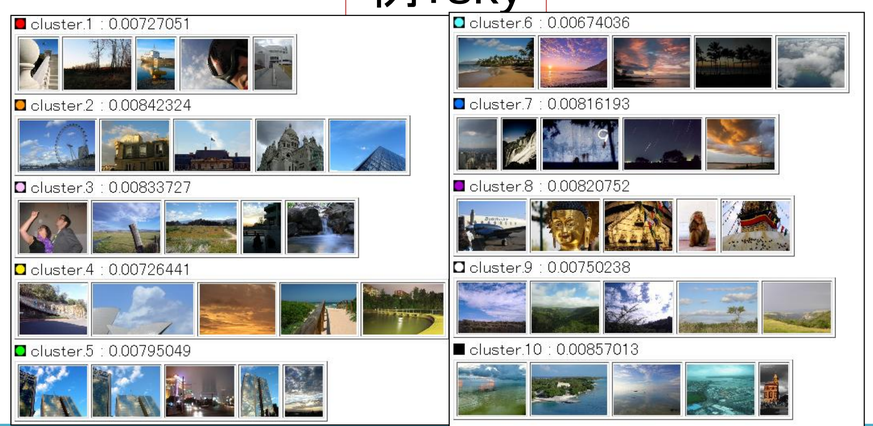
指標値の大きい10語		指標値の小さい10語	
指標値	単語	指標値	単語
0.00940416	building	0.00605590	california
0.00935062	holy	0.00613669	deutschland
0.00924220	airplane	0.00633069	machu picchu
0.00918682	beach	0.00675465	niagara
0.00912758	half	0.00706835	uluru
0.00911686	field	0.00735098	arc de triomphe
0.00908912	monkey	0.00749106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



例：doragon

例：building

例：sky



まとめ

- ◆単語概念の視覚の地域性を定量化する手法を提案
- ◆ほぼ同一の複数の画像で学習した場合、影響が生じる

↓

- ◆Flickrの投稿IDや位置情報、撮影日時から重複・類似画像を省く必要