

VisualRankにおける位置情報 活用の検討

電気通信大学大学院

情報工学専攻

川久保 秀敏 柳井 啓司

概要

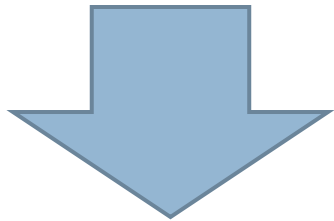
- 本研究では, VisualRankを用いた位置情報付き画像のランキング手法を提案
- 提案手法では, 位置情報付き画像の視覚特徴と位置情報の両方を考慮してランク付けを行なう
- 提案手法で, 上位にランキングされる画像は
 - ▣ 代表的な視覚特徴をもつ画像
 - ▣ パラメータとして与えた座標に, 近い地点で撮影された画像

概要

- 視覚特徴
 - カラーヒストグラム
 - SIFT記述子による Bag-of-Features表現
- 350語の単語について, Flickrから位置情報付き画像を収集し, 実験を行った

研究の目的

- 提案手法を用いて、
視覚特徴と位置情報の両方を考慮した
ランキングを得る



- クエリに関する、地域ごとの代表画像を得る

各地の代表画像の例

結果例: house に関する, 各地の代表画像

アジア

座標パラメータ: 東京

アジア諸国の家



アメリカ

同: ニューヨーク

洋風の家
ホワイトハウス



オーストラリア

同: シドニー

洋風の家
オペラハウス



地域別代表画像の利用

単語に関する, 地域ごとの代表画像を得る

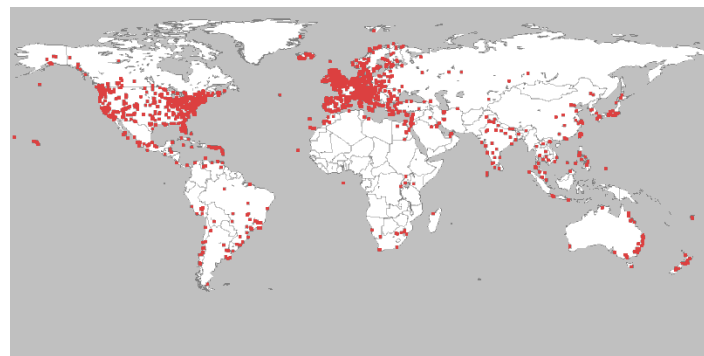
関連研究

- 川久保秀敏, 柳井啓司, “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”, CVIM, 2009, 3月
- 位置情報を一般物体認識に活用することを目指す研究
- 単語の視覚性と地域性を自動的に計算するために, 単語に関する収集画像を分析.

例:sun



視覚のエントロピー 小



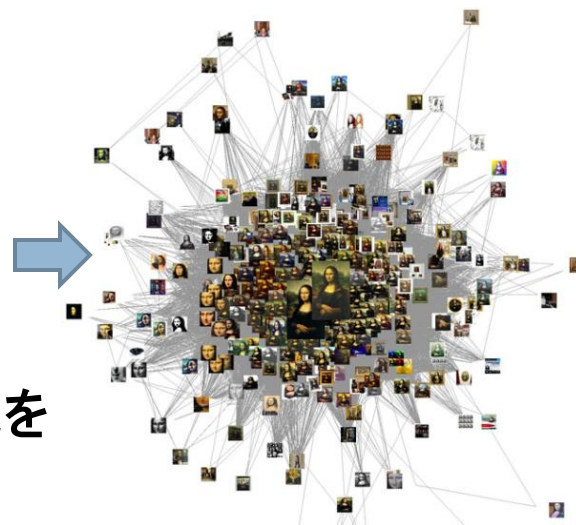
位置のエントロピー 大

利用手法 VisualRank

- VisualRank
 - ▣ PageRankアルゴリズムを画像に適用したもの
 - ▣ Jingらによって提案された
 - Y.Jing and S.Baluja, "VisualRank: Applying pagerank to large-scale image search," IEEE Trans PAMI, 2008
- 画像間の類似度をもとに、画像のランキングを得る。

モナリザ画像の
類似関係のグラフ

中心に、代表的な視覚を
持つ画像

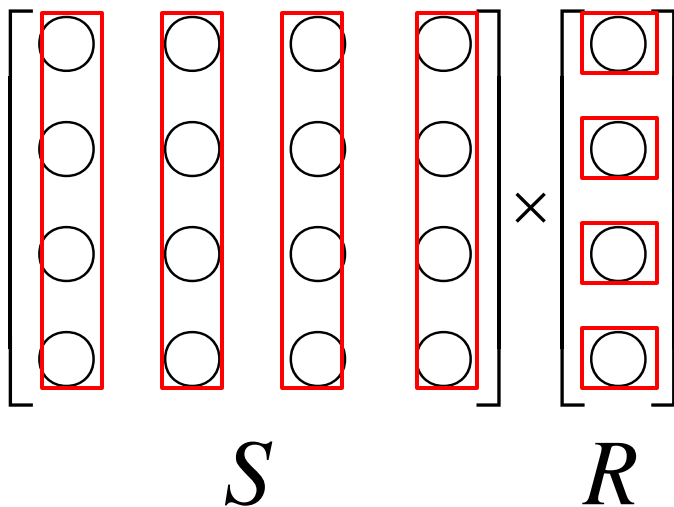


VisualRank

□ 更新式(補正ベクトルなし)

$$R = S \times R \quad \text{VisualRank値ベクトル}R\text{を}S\text{にかけることで更新}$$

□ 更新式を反復し, Visualrank値を収束させる



R : 各画像のVisualRank値が
並んだベクトル

S : 画像の類似度行列を
列ごとに正規化したもの
(正規化することで, 更新時に
VisualRank値の合計が変化しない)

VisualRank

□ 更新式(補正ベクトルあり)

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

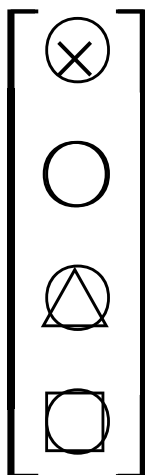
$$\alpha \begin{bmatrix} \bigcirc & \bigcirc & \bigcirc & \bigcirc \\ \bigcirc & \bigcirc & \bigcirc & \bigcirc \\ \bigcirc & \bigcirc & \bigcirc & \bigcirc \\ \bigcirc & \bigcirc & \bigcirc & \bigcirc \end{bmatrix} \times \begin{bmatrix} \bigcirc \\ \bigcirc \\ \bigcirc \\ \bigcirc \end{bmatrix} + (1 - \alpha) \begin{bmatrix} \bigcirc \\ \bigcirc \\ \bigcirc \\ \bigcirc \end{bmatrix}$$

S R P

VisualRank

□ 更新式(補正ベクトルあり)

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



P

P : 補正ベクトル

・均一な場合

VisualRank値の均等化

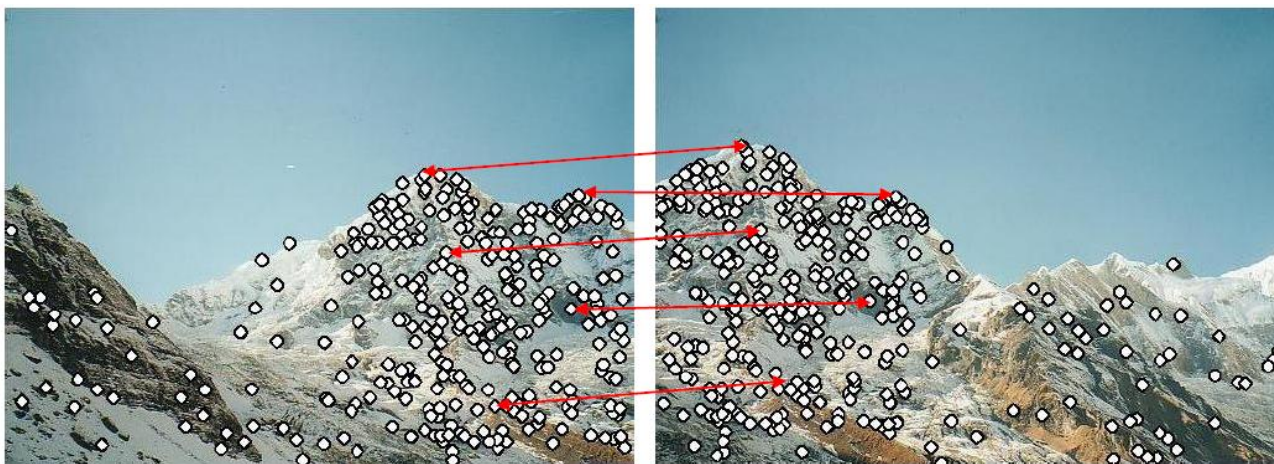
・不均一な場合

一部の画像を強調するバイアス

α : 補正の強さを調整するパラメータ
0.8~0.9程度の値が良いとされる

SIFT対応点による類似度

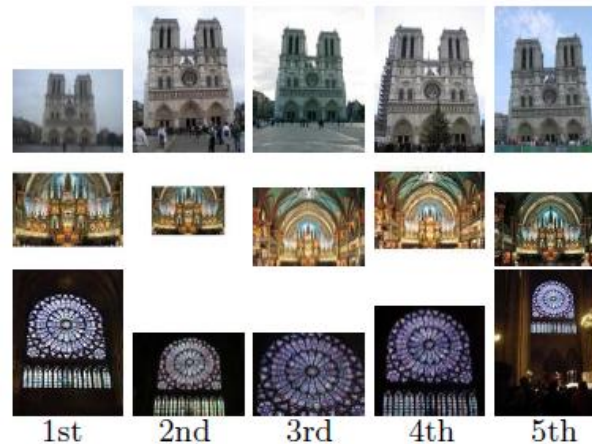
- Jingらは類似度にSIFT特徴の対応点数を利用



- 商品画像やランドマーク画像など、同一物体が写っていることが予想される場合に有効
- 上位に同一物体の画像が並びやすい

関連研究: MultiClass VisualRank

- 安倍らによって提案された、多様な上位画像を並列に提示する手法
 - 安倍満, 吉田悠一, “Visualrank の多クラスへの拡張: 画像特徴量を用いた類似画像の自動分類とランキング付け手法” PRMU2008-178 · pp.183-188
- まず画像のクラスタリングを行い, 各クラスタについてそれぞれVisualRankを適用する.
- 各クラスタの上位画像を並列に提示して, 多様性を確保.



← クラスタ1のランキング

← クラスタ2のランキング

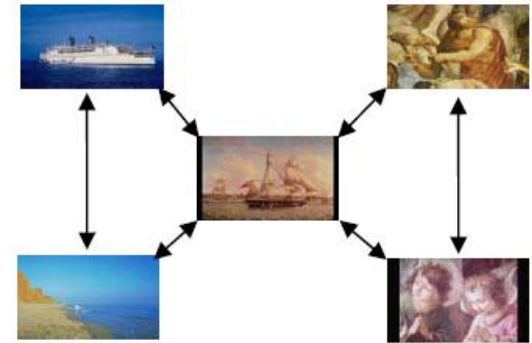
← クラスタ3のランキング

関連研究： ImageRank

□ Heらによる, 画像のランク付けに関する研究

- X.He, W.Y.Ma, and H.Zhang, "ImageRank: spectral techniques for structural analysis of image database," IEEE ICME, 2003.

□ 画像の近傍関係をグラフ化



- マルコフモデルによる画像間の遷移確率を求める
- 遷移確率行列の固有ベクトルを用いて代表画像を決定.

提案手法

- VisualRankを用いて位置情報付き画像に対するランク付けを行う。
- 位置情報付き画像の視覚特徴と位置情報の両方を考慮したランキング
 - パラメータ α によって, どちらを重視するかを設定

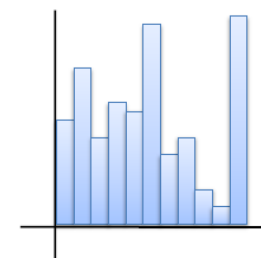
提案手法

- 視覚特徴には、カラーヒストグラムとSIFT記述子によるBoF表現を使用。
 - 様々な単語をクエリとするため、一般物体認識でよく用いる特徴を利用

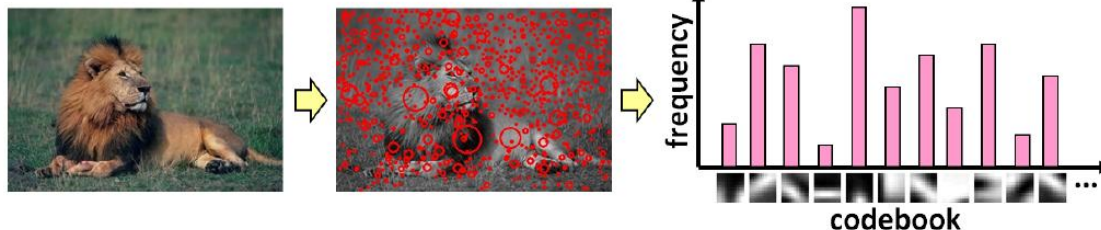
類似度

□ 視覚特徴量

- 64次元のカラーヒストグラム（RGB色空間）



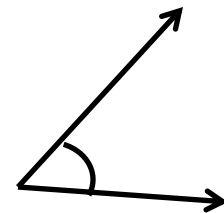
- 500次元のBag-of-Features表現（SIFT記述子）



- 類似度には、ヒストグラムインタセクションを使用
- 各特徴量で類似度行列を求めておき、合成

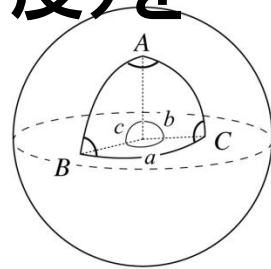
類似度

- 視覚特徴の代わりにテキストタグを用いる場合
 - 画像に多く付与された上位500のタグを求める
(クエリごとに求める, クエリ自体は除く)
 - 上位タグについてのバイナリベクトルを
画像ごとに作成
(0:付与されていない, 1:付与されている)
 - 画像間の類似度は, タグバイナリベクトルの
コサイン類似度



位置情報によるバイアス

- 注目点に近い座標で撮影された画像に大きい値を与える
- 各画像の撮影位置情報と注目点の距離(角度)を球面三角法で求める.
 - 地球を半径1の球とする



$$D_i = \cos^{-1}(\sin(lat_i)\sin(lat_A) + \cos(lat_i)\cos(lat_A)\cos(long_i - long_A))$$

$lat_i, long_i$: 画像 i の位置情報

$lat_A, long_A$: 注目点の座標

位置情報によるバイアス

- 注目点に近い座標で撮影された画像に大きい値を与える

論文中の式(4)の訂正

正
$$p_i^{geo} = \frac{\pi - D_i}{\pi}$$

誤
$$p_i^{geo} = \frac{D_i}{\pi}$$

位置情報によるバイアス

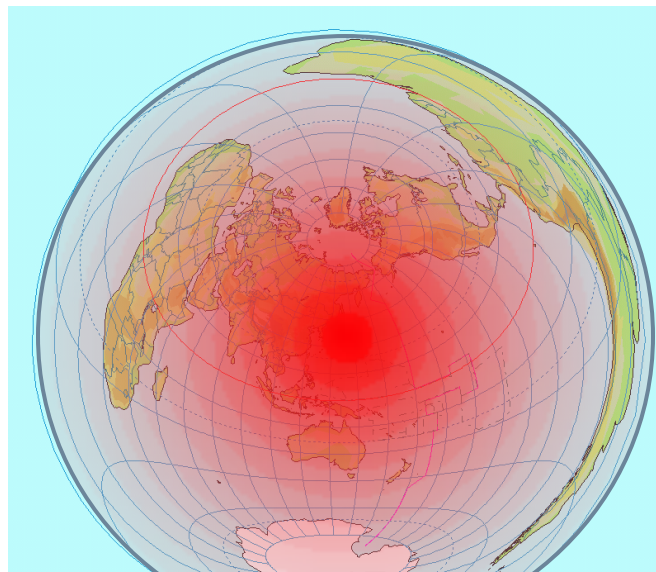
- 注目点に近い座標で撮影された画像に大きい値を与える

$$P_i^{geo} = \frac{\pi - D_i}{\pi}$$

P_i^{geo} はVisualRank値の合計値に正規化される。

$$P_i^{geo} = \frac{P_i^{geo}}{\|P_i^{geo}\|_1} \|R\|_1$$

東京を注目点とした場合



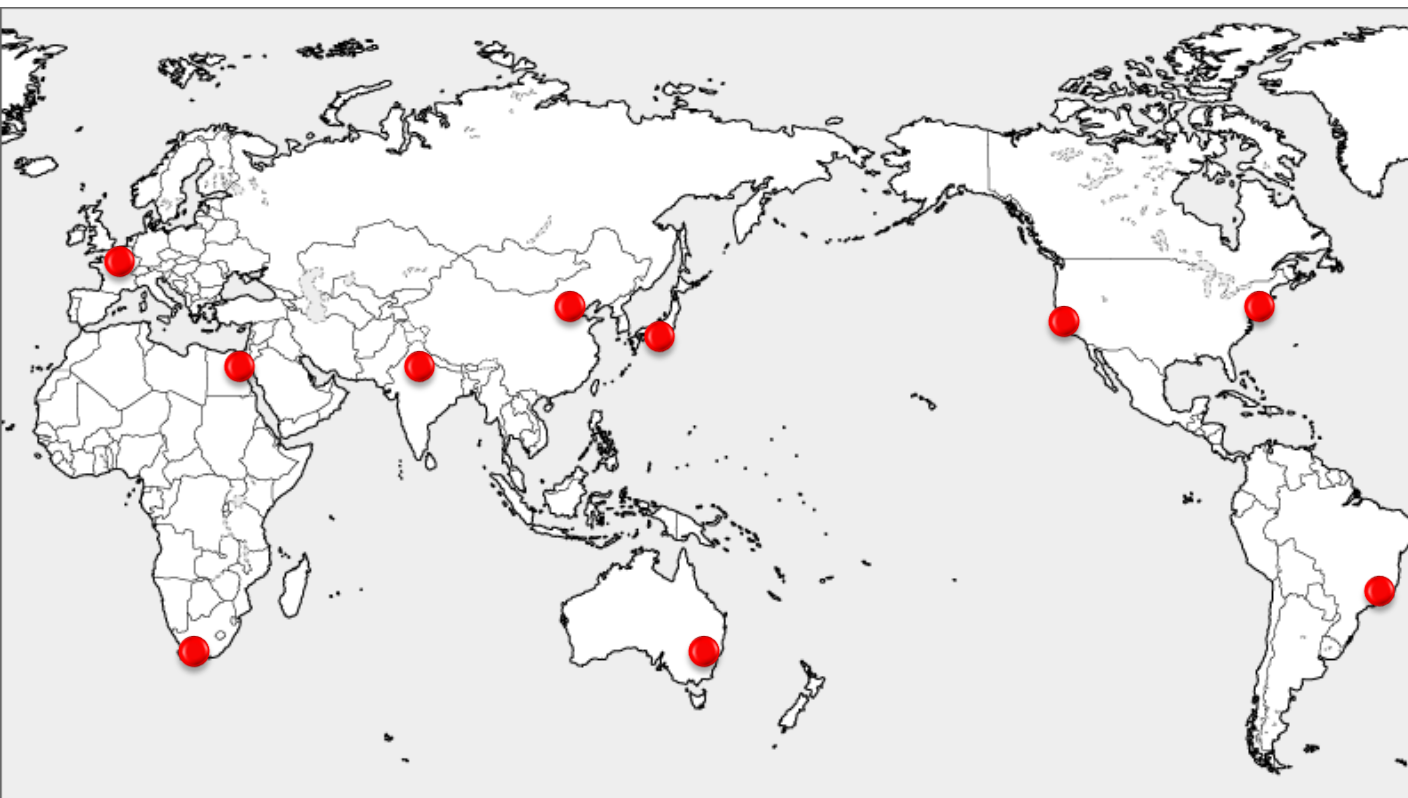
東京と中心とするアジアの画像に大きい値

実験

- 名詞250語と形容詞100語をクエリ単語とした。
- Flickrから位置情報付き画像を収集
 - 各単語について2000枚ずつ収集
 - 同一ユーザが投稿した画像は10枚まで
- 各単語について、提案手法でランキングを求めた。

実験

- 注目点として、以下の10都市の座標を用いた。



東京

北京

シドニー

サンフランシスコ

ニューヨーク

リオデジャネイロ

パリ

カイロ

ケープタウン

デリー

結果

□ 実験結果を

<http://mm.cs.uec.ac.jp/kawaku/geovisualrank/>
で公開中

□ サーバーが不調な場合があり，原因を調査中



The screenshot displays the geovisualrank web application interface. At the top, there are navigation tabs for '地図' (Map), '結果写真' (Result Photos), and '地図+写真' (Map+Photos). The main area shows a world map with various colored markers (red, green, blue, yellow) indicating different data points across various countries. A pop-up window is open over a location in Europe, displaying the following information:

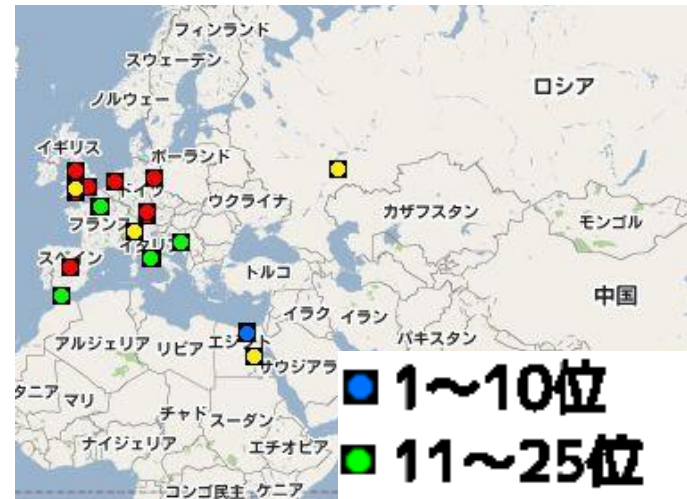
- 46番目の画像 (46th image)
- visualrank:1.164896
- 緯度:55.745261 (Latitude: 55.745261)
- 経度:37.606276 (Longitude: 37.606276)

Below the map, there are several control panels:

- 名詞 (Nouns):** A list of words including snake, socks, son, pound, spider, sport, square, starfish, statue_of_liberty, and steak.
- 形容詞 (Adjectives):** A list of words including aerial, ancient, enbuse, bad, beautiful, best, better, big, black, and blue.
- 都市 (City):** A dropdown menu for selecting a city.
- 表示数 (Number of results):** A dropdown menu set to 100.
- feature:** A dropdown menu set to color,BoF 各50%.
- alpha:** A dropdown menu set to 0.90.
- 現在の状態 (Current state):** A section with input fields for '単語:' (set to alexander), '都市:' (set to tokyo), '表示数:' (set to 100), '特徴量:' (set to color,BoF 各50%), and 'バイスパラメータα:' (set to 0.90).

結果 各地のpyramid

注目点 : カイロ $\alpha : 0.85$
視覚特徴: color, BoF 50% ずつ



注目点 : パリ $\alpha : 0.85$
視覚特徴: color, BoF 50% ずつ



結果 各地のpyramid

メキシコのピラミッドが上位に。
ブラジルのピラミッド画像が殆どなかったため

対応策

- 上位画像と注目点の距離で、結果の信頼度を評価
- 上位画像の平均位置を新注目点として再度ランキング

注目点 : リオ $\alpha : 0.85$
視覚特徴: color, BoF 50%



結果 traditional

- 文化に関する単語は地域差が出やすい？
- 各地の traditional $\alpha:0.85$ 特徴: color, BoF 50%

注目点: シドニー



注目点: 東京



注目点: リオ



結果 game

- game α : 0.85 特徴: color, BoF 50%

注目点: 東京

テレビゲーム関係の画像



注目点: ケープタウン

屋外の画像



結果 特徴量による変化

□ 注目点パリ $\alpha = 0.85$

カラーヒストグラムのみ



SIFTによるBoF表現のみ



テキストタグ



上記3種を33%ずつ



結果 特徴量による変化

- 各特徴をそれぞれ用いた結果をまとめる方が多様な画像をもつデータセット作成に生かせると予想

カラーヒストグラムのみ



SIFTによるBoF表現のみ

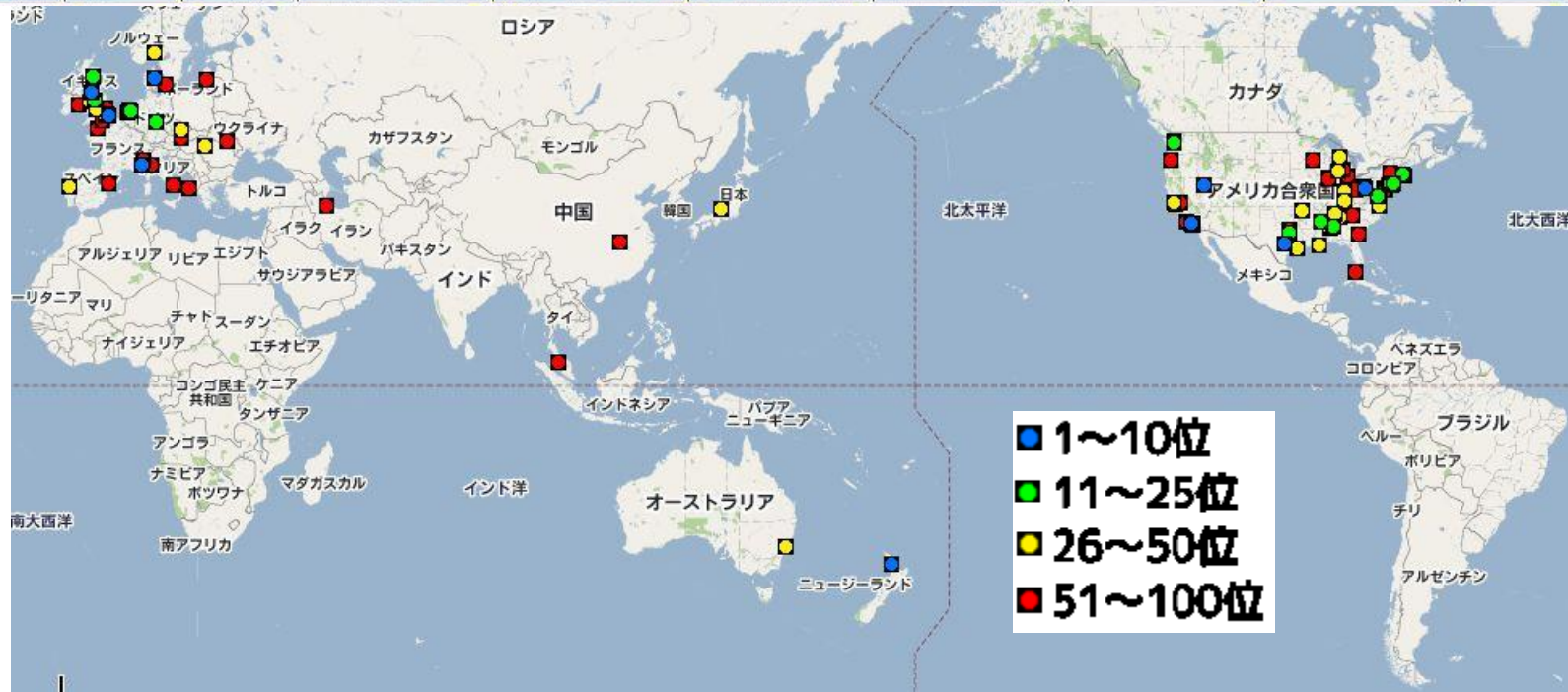


上記2種を50%ずつ



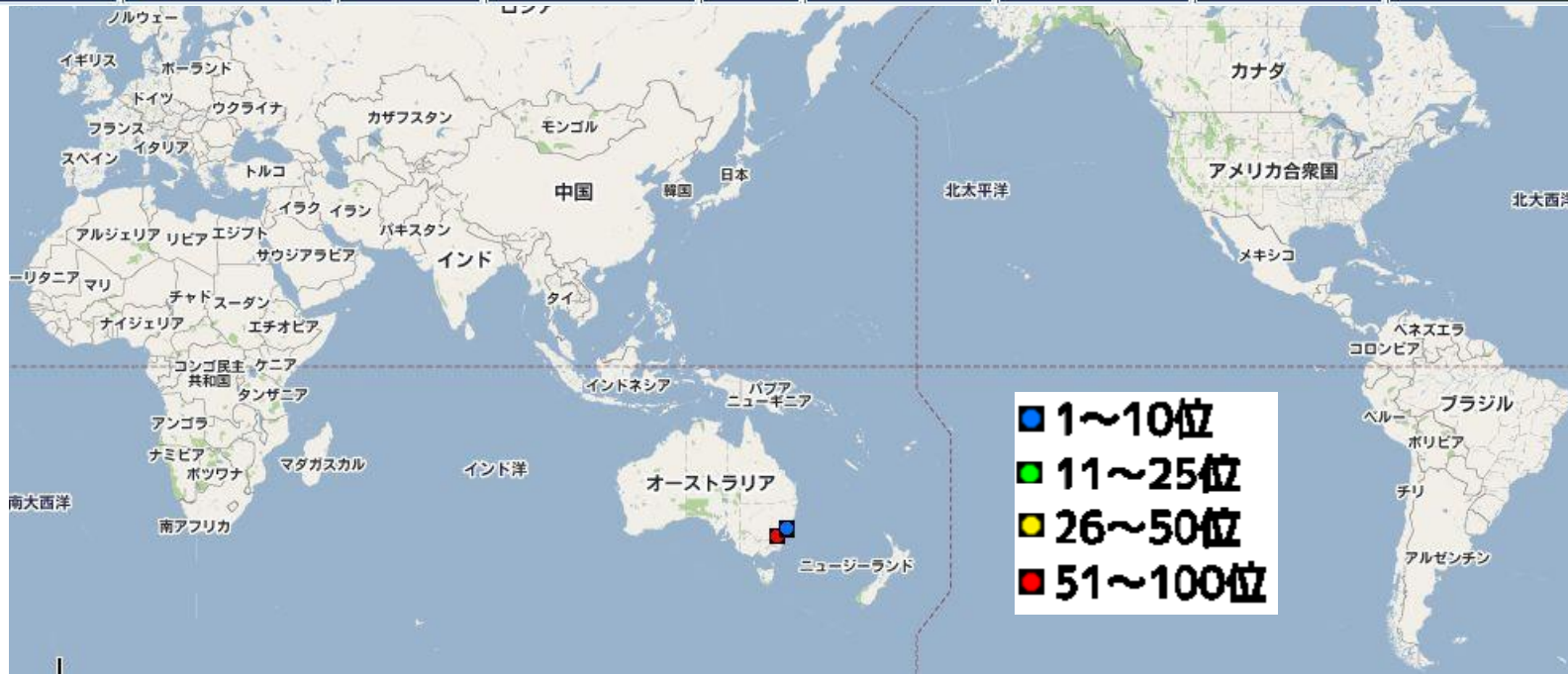
結果 house aによる変化

- “house”での結果 バイアスなし, $\alpha=1.0$
- 特徴量は color 50%, BoF 50%



結果 house aによる変化

- $a=0.0$, 注目点に近い順



結果 house aによる変化

□ $a=0.5$



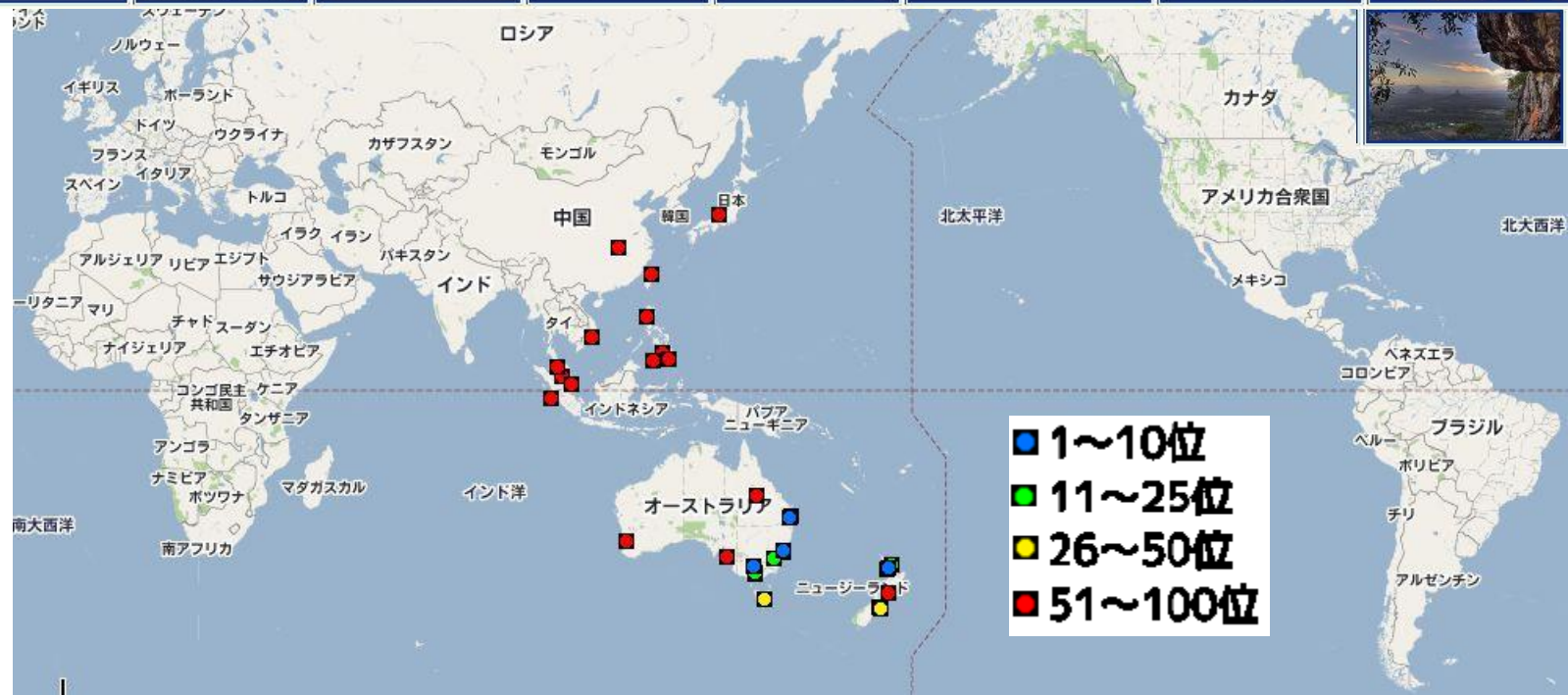
結果 house aによる変化

□ $a=0.8$



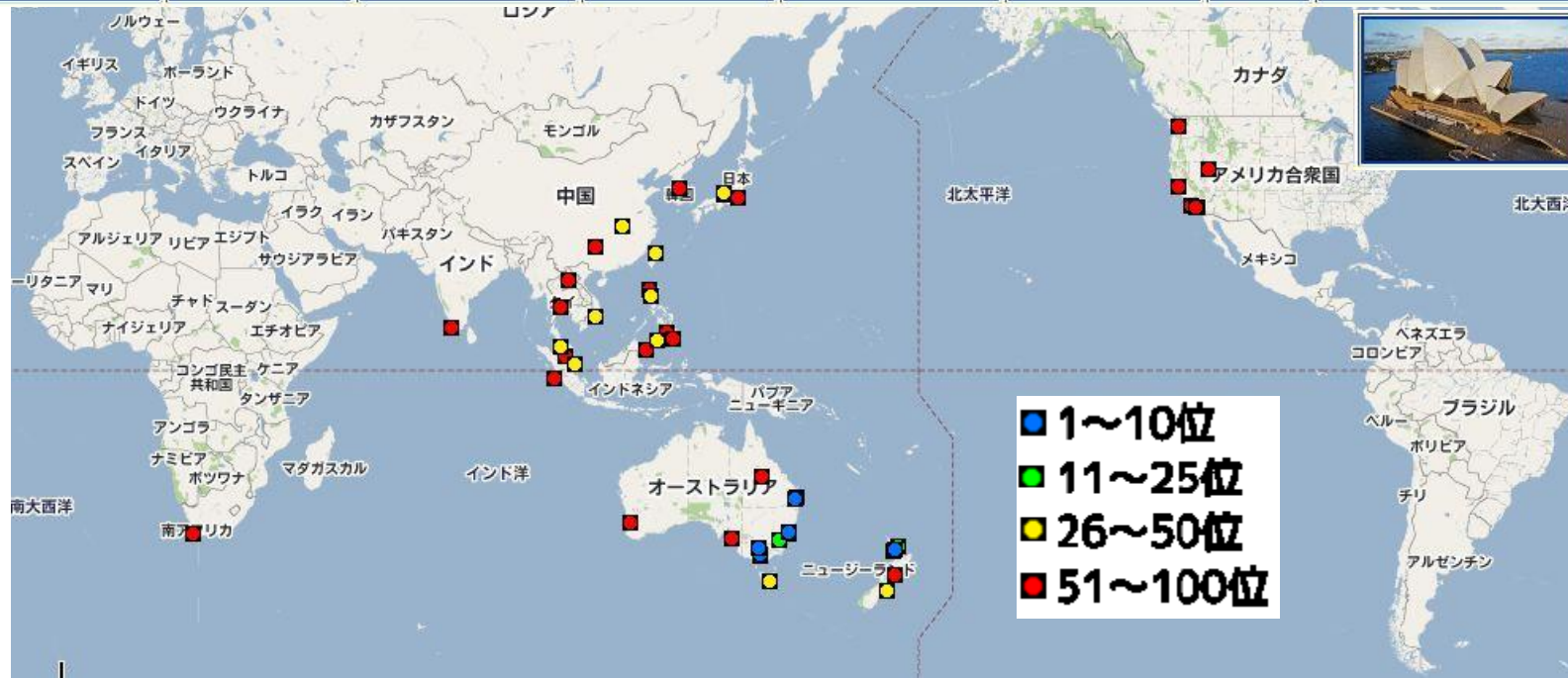
結果 house aによる変化

□ $\alpha=0.85$



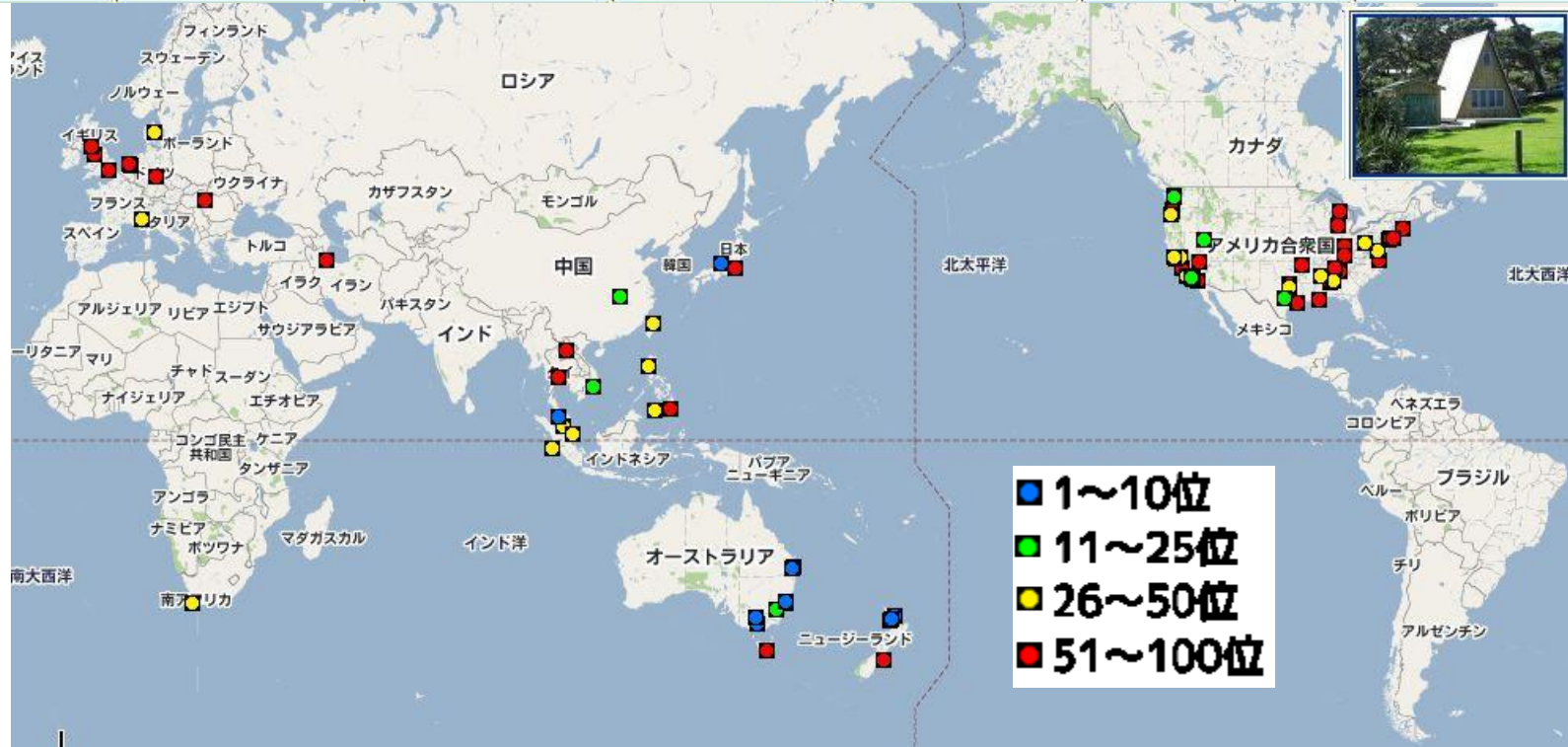
結果 house aによる変化

□ $\alpha=0.90$



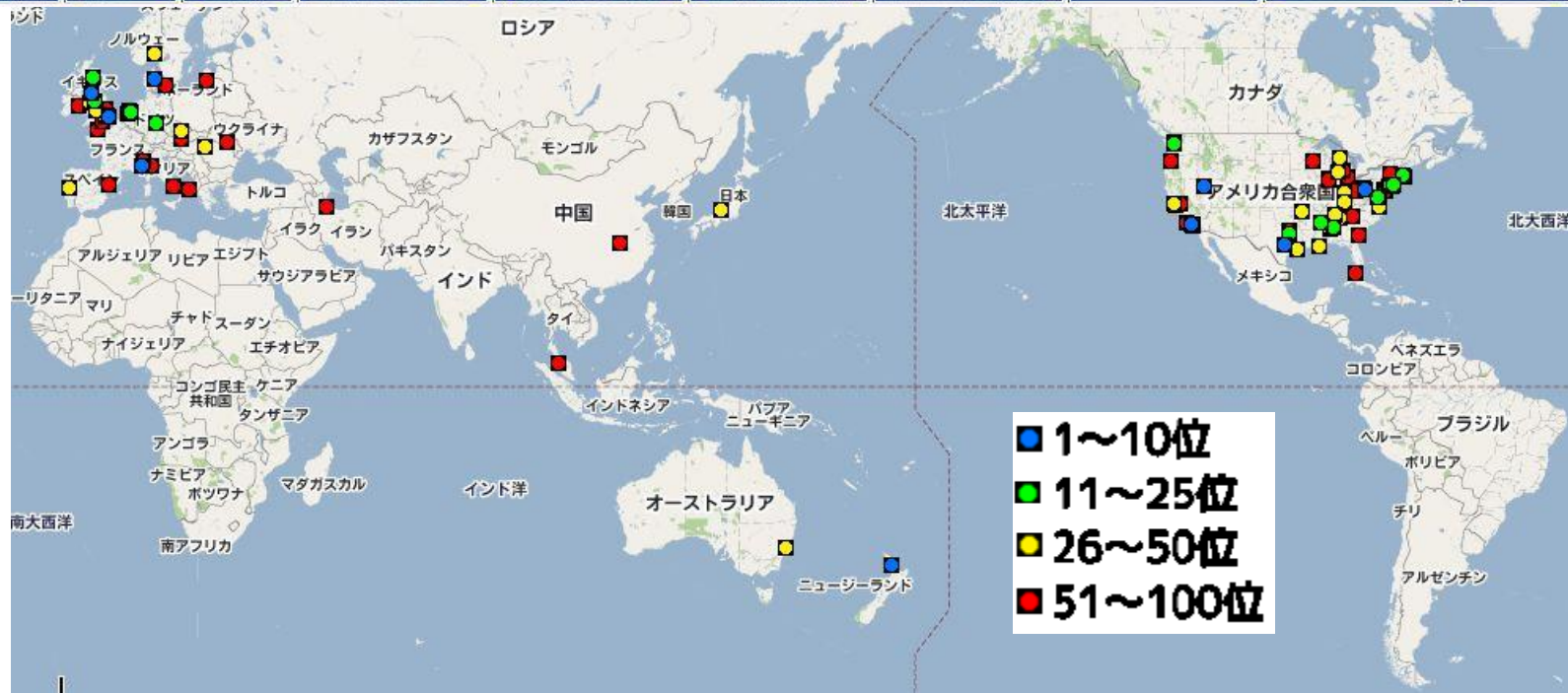
結果 house aによる変化

□ $\alpha=0.95$



結果 house aによる変化

- “house”での結果 バイアスなし, $\alpha=1.0$
- 特徴量は color 50%, BoF 50%



今後

- 提案手法の評価
 - ▣ 被験者に評価させる(Amazon Mechanical Turk ?)
- 他手法との比較
 - ▣ 位置情報での絞り込んだ画像でのVisualRank
 - ▣ テキスト検索時のクエリに地域名を足した場合
 - ▣ 位置情報によるバイアスの作成方法を変えたもの
- 評価・比較をもとに、提案手法の改良を行ない、位置情報付き画像データセットの自動作成へ活用

まとめ

- VisualRankを用いた位置情報付き画像のランキング手法を提案
- 視覚特徴と位置情報の両方を考慮したランキング
- 提案手法を用いて, 350単語に関するFlickr画像のランキングを生成した.
- 今後, 提案手法の評価・改良・活用