



**Web 上のジオタグ画像を用いた
世界各地の文化的差異の発見**

**2009年度 人工知能学会 全国大会
2009年6月 高松**

**柳井啓司
電気通信大学 情報工学科**

研究の背景



Webには、ラーメンがいっぱい

やっぱり、 どこのもラーメンが知りたい！



アウトライン

- **研究の背景・目的**
- **関連研究**
- **方法**
- **実験結果**
- **まとめと今後の課題**

背景: 大量のジオタグ画像の登場

- 近年, 位置情報付き画像 (**geo-tagged image**) が急速にWeb上に増えている. **Flickr, Panoramio**
 - Flickrは 4千万枚の位置情報付き画像を保有



(in the beginning of 2008)

「ジオタグ画像」
(撮影)位置の情報
(緯度, 経度)を
メタデータとして
持っている画像.

Geotagged images

研究の動機 : *Objects over the World*

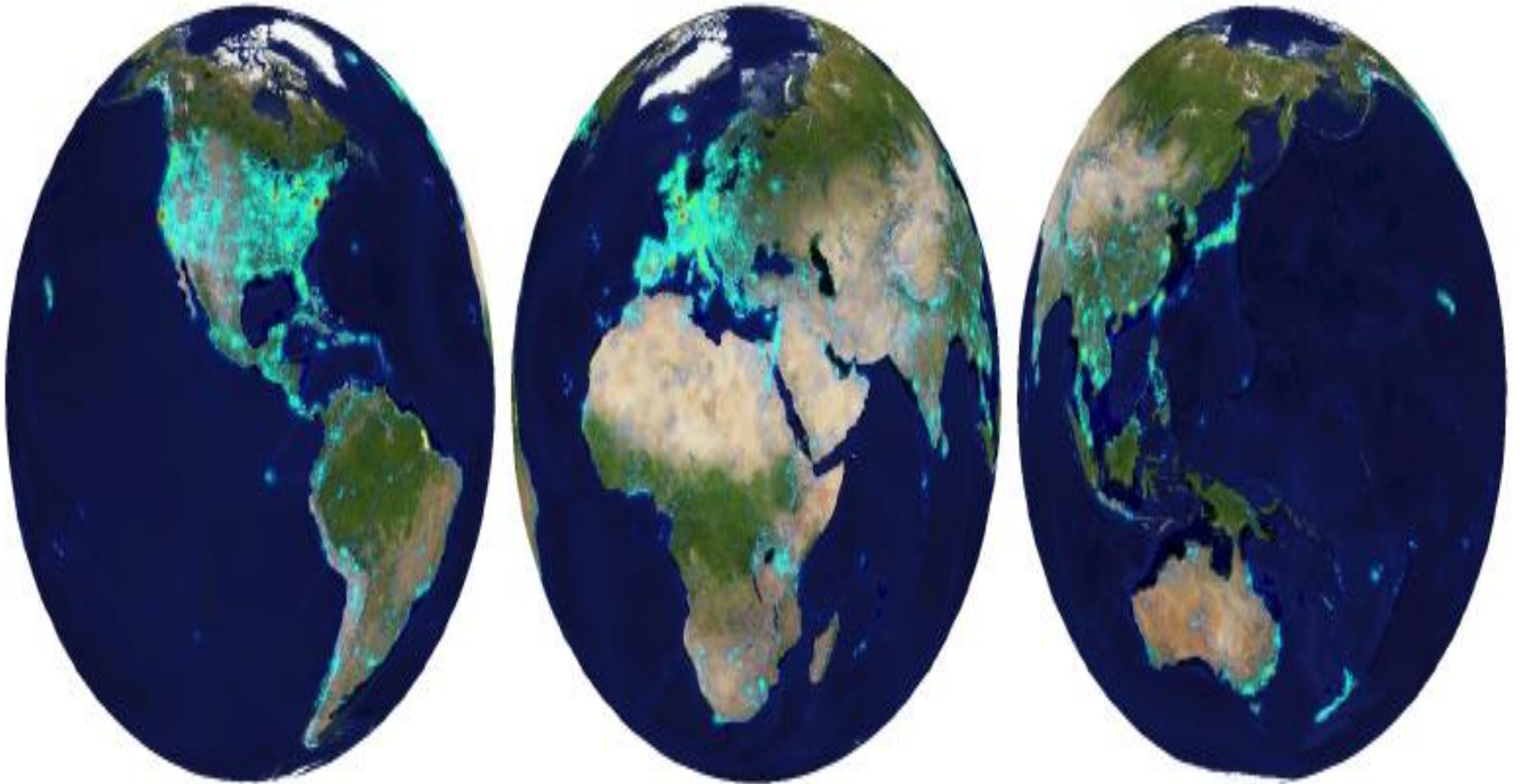
■ 大量のジオタグ画像を使って、 世界中の“X”画像を集めてみよう！

- ✓ Do you know all kinds of famous “noodles” in the world?
 - ✓ “Ramen” and “Soba” in Japan, “Thai noodle” in Thailand, “Chinese noodles”, “rice noodle” Taiwan, “Spaghetti” in Italy...



- ✓ How do such scenes as “beach”, “waterfall”, “mountain” look like in different areas in the world?
- ✓ How about other objects such as “flower”, “castle”, “clothes”, “car”
→ 写真じゃないと分からない！

ジオタグ画像の分布



James Hays and Alexei A. Efros, Carnegie Mellon University: IM2GPS: estimating geographic information from a single image, CVPR, (2008)



Share This

Add to your map

You're looking at **Tokyo Prefecture, Japan**. Try dragging the map around or searching.

GO

(Some formats are OK.)

SAVE TO MAP Or, cancel.

35°42'39"N, 139°17'05"E
(35.7108, 139.2846)

Go to the Organizer | Go to the map

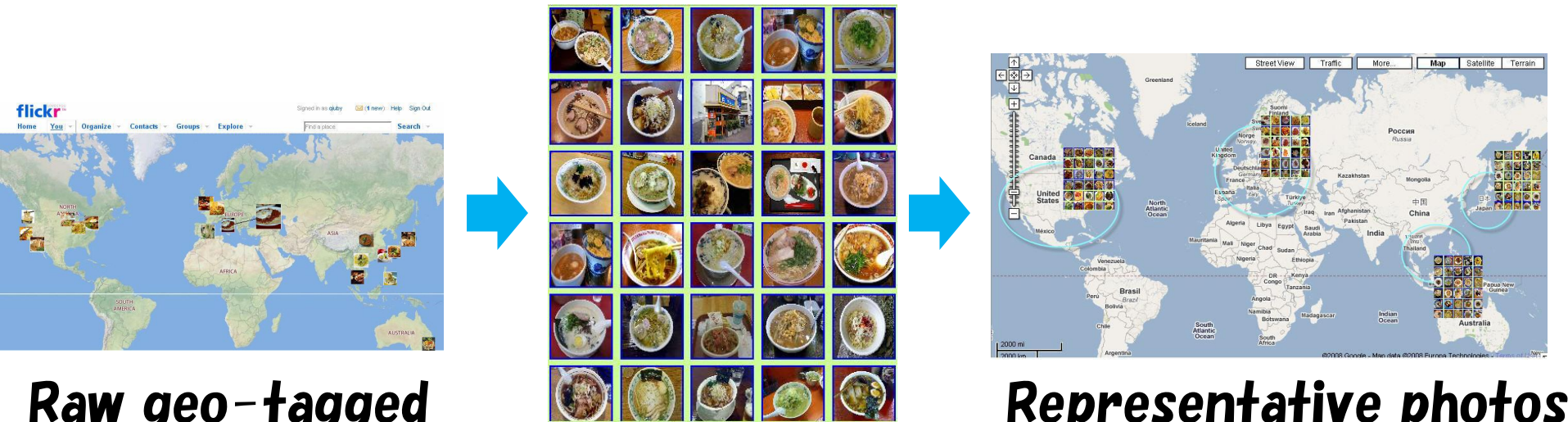
©2008 NAVTEQ ©2008 Yahoo! Inc.

- **2006年 Flickrが geotagging UI を公開**
 - **爆発的增加. 4千万枚. 10万枚/月増加中.**

目的

Flickrのジオタグ画像を利用

特定カテゴリー”X”画像が (e.g. X=“ラーメン”) 存在する世界中の代表的な地域と、その地域で代表的な画像を Web上のジオタグ画像を用いて自動抽出。



Raw geo-tagged photos on Flickr
Most relevant photos selected by our system

Representative photos generated for typical regions

関連研究

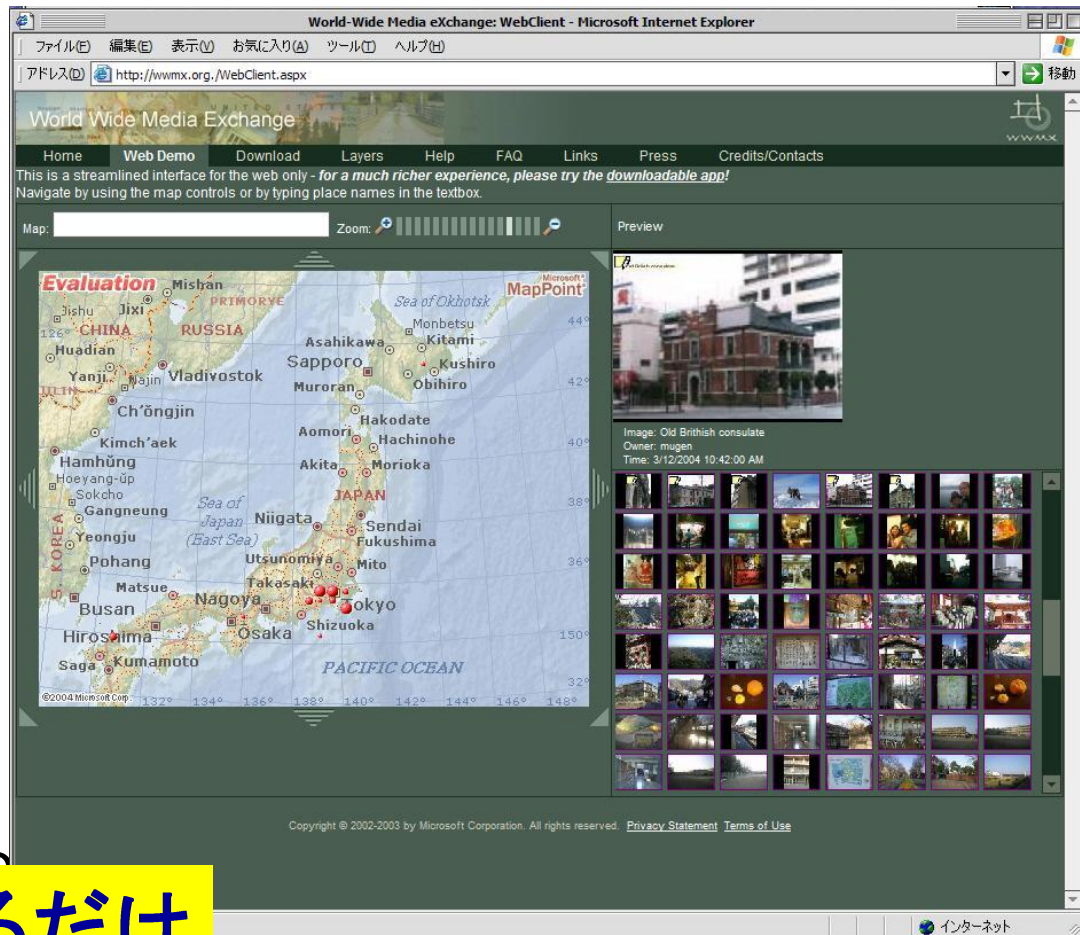
【位置情報付き画像ミニサーベイ】

位置情報付き画像の研究：初期

■ WWMX (Toyama et al., 2003)

■ MS Researchの Webベース 位置情報画像 共有システム

(Flickrは2004～)



K. Toyama, B. Logan, A. Bo

ただ地図上に表示するだけ

ACM MM, pp. 100-100, 2003.

画像のグループ化とタグ付け



M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina:
"Automatic Organization for Digital Photographs with
Geographic Coordinates", ACM MM, pp.53-62, 2004.

関連研究: WorldExplorer

■ 位置に基づくタグのクラスタリング

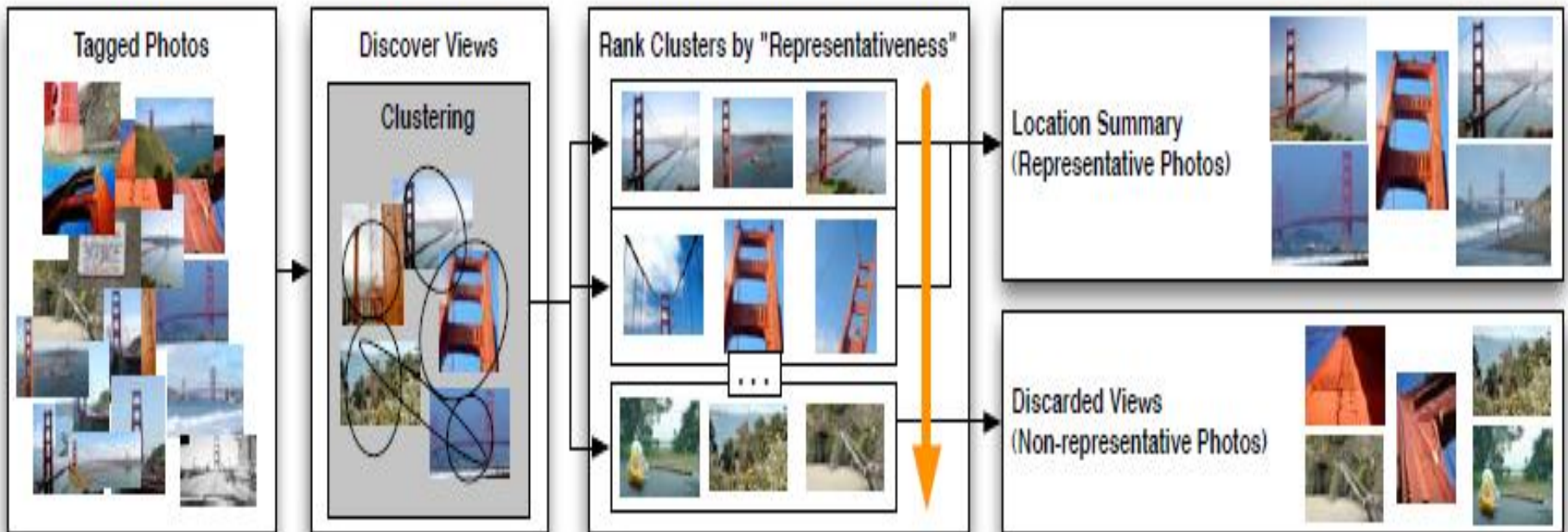


Ahern, S. , Naaman, M., Nair, R. and Yang, J. : World Explorer: Visualizing

Flickrタグを位置でクラスタリング(位置画像のタグ)

関連研究：画像特徴を利用

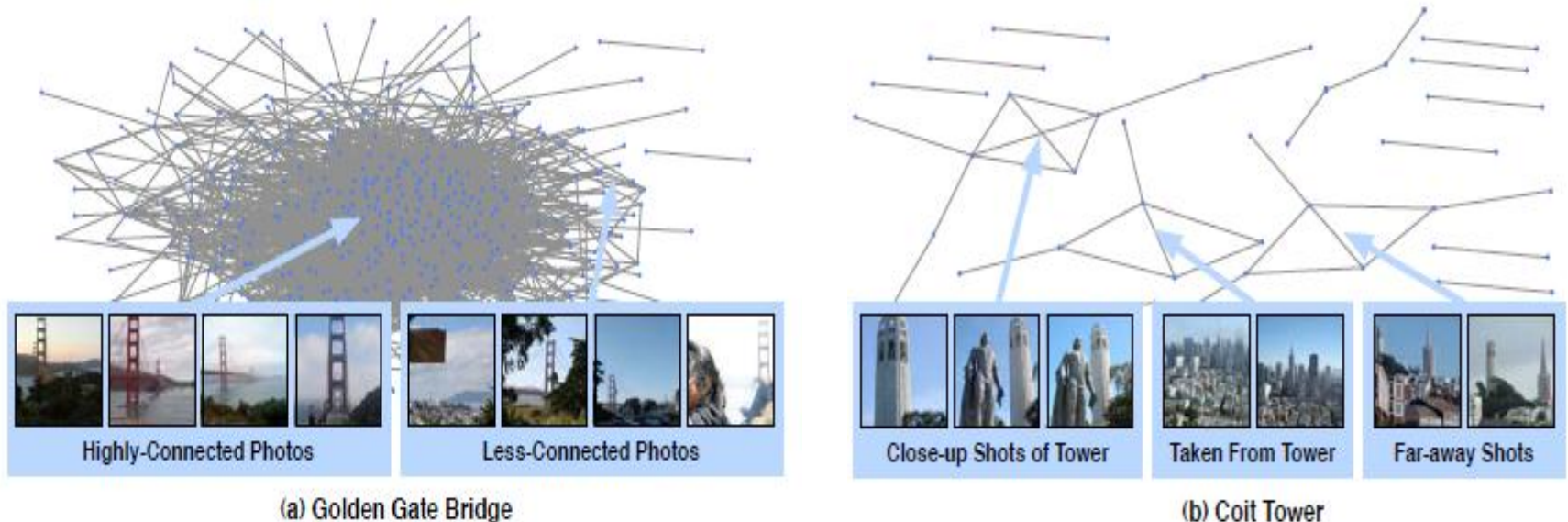
■ ランドマークの代表画像を自動選択.



Lyndon Kennedy and Mor Naaman: Generating Diverse and Representative Image Search Results for Landmarks, ACM WWW2008, pp.297-306, (2008).

関連研究：画像特徴を利用（続）

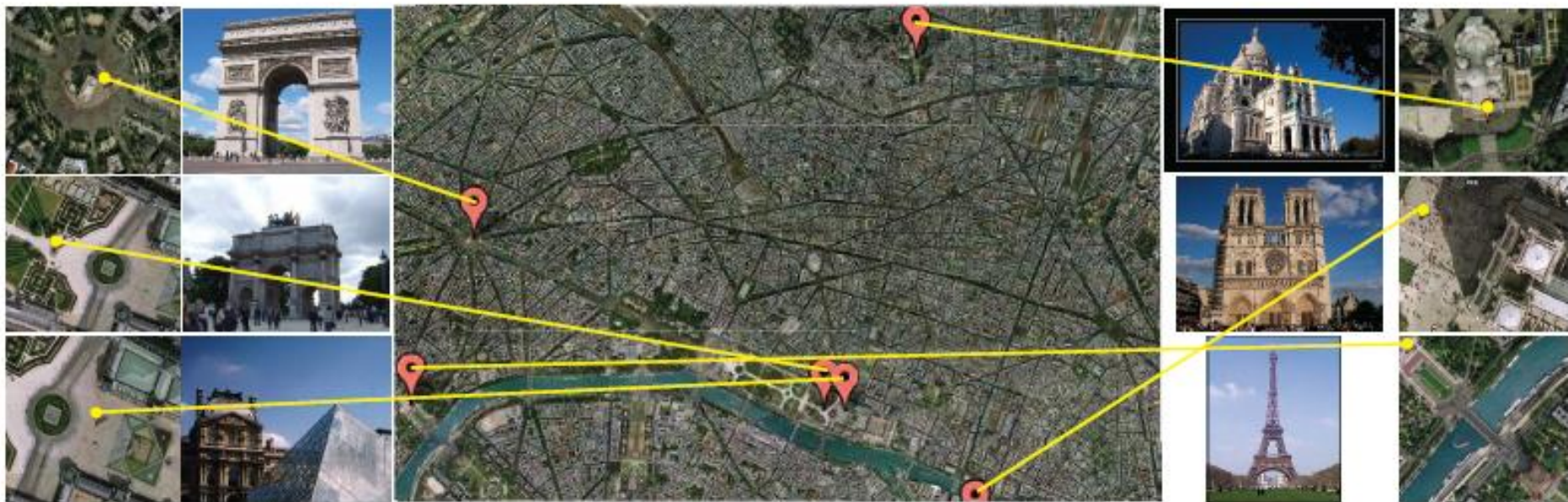
- ランドマークの代表画像を自動選択.
- **ランドマーク**はどの写真も同一対象を写しているので, SIFT matching によるグラフのクラスタリング



SIFT-based の類似度を用いた代表画像抽出

指定範囲の代表画像の抽出

- 指定範囲(100m四方の領域)の代表的画像(おもにランドマーク)抽出

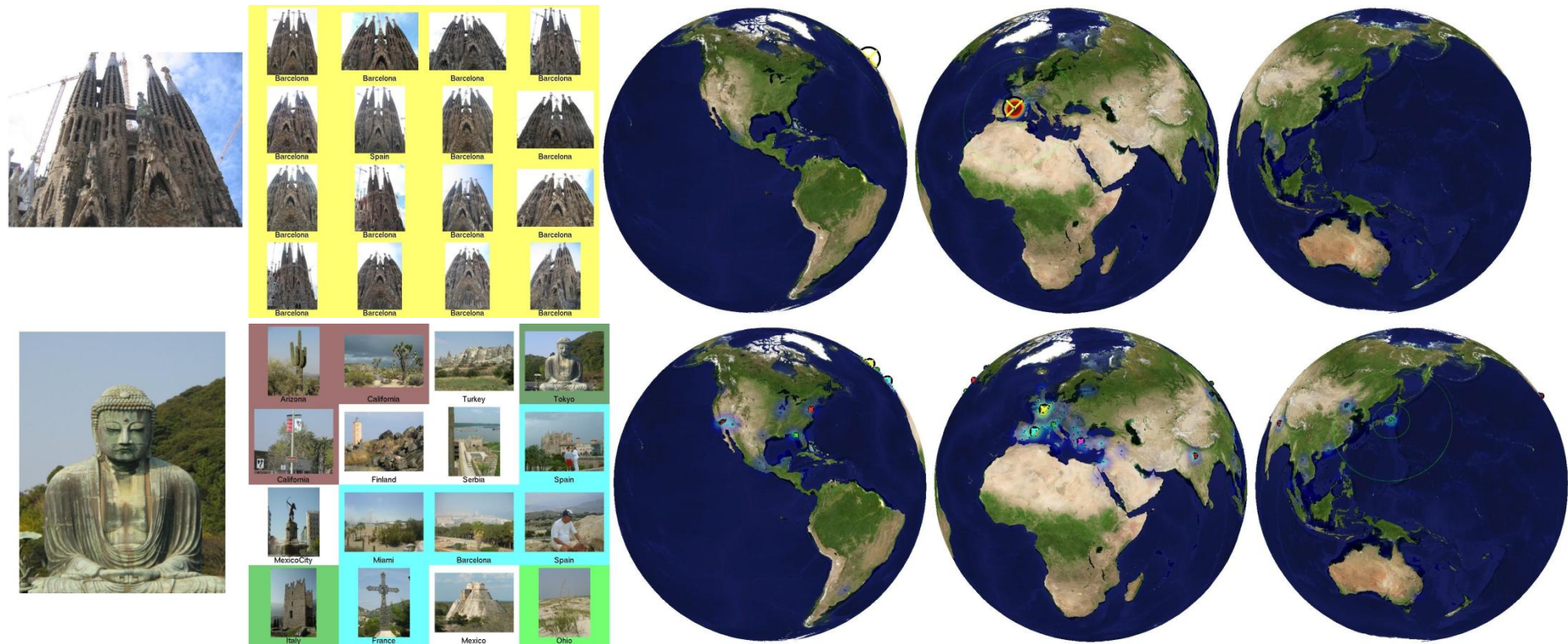


Till Quack, Bastian Leibe and Luc Van Gool: World-scale Mining of Objects and Events from Community Photo Collections. ACM CIVR, pp. 47-56

指定範囲の代表ランドマーク画像. これもSIFT-based.

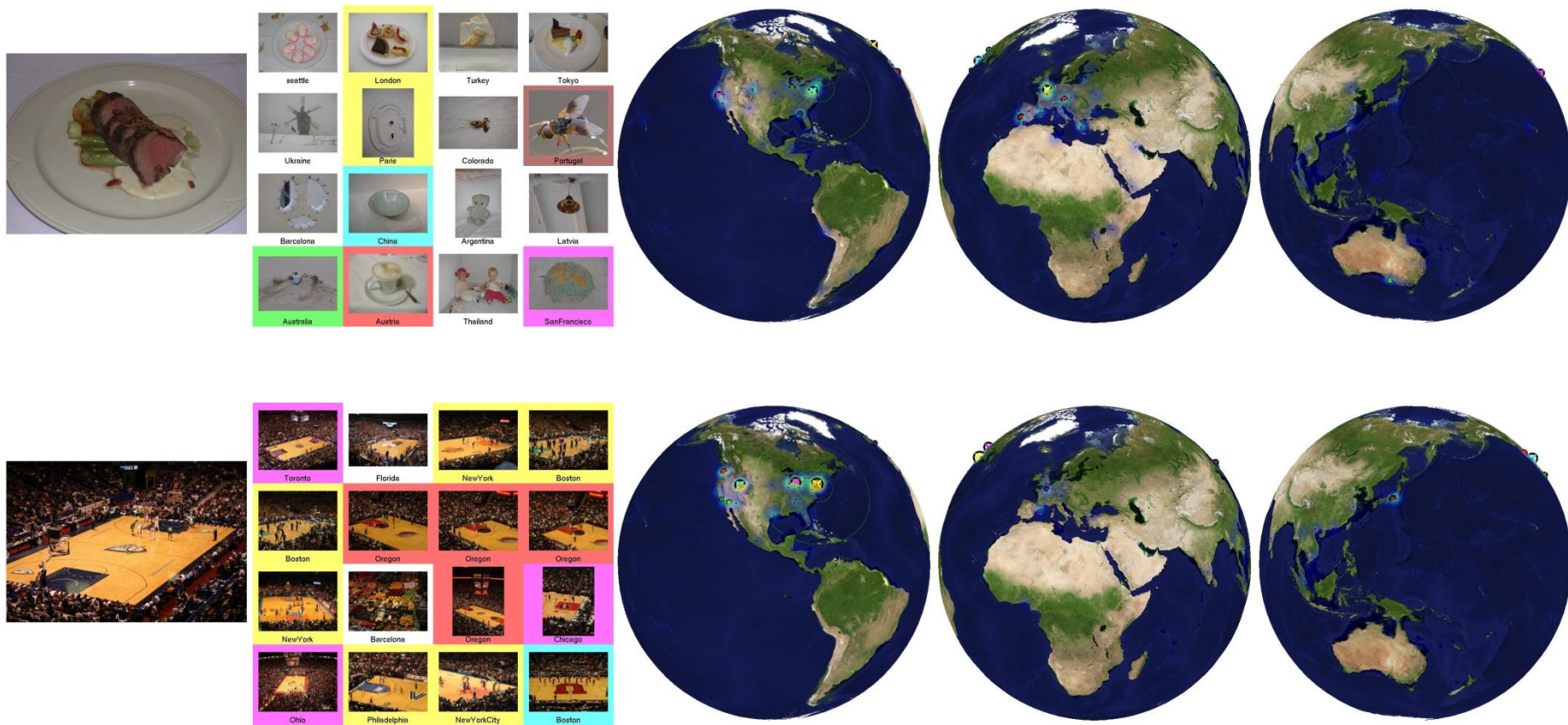
600万枚のジオタグ画像に 対するNNで、撮影位置を推定

■ *Brute-force object recognition* の一種



James Hays and Alexei A. Efros: IM2GPS: estimating geographic information from a single image, CVPR, (2008).

600万枚のジオタグ画像に対するNNで、撮影位置を推定 (続) 18



James Hays and Alexei A. Efros: IM2GPS: estimating geographic

大量のジオタグ画像を利用し、NNで認識.

対応する位置の航空写真と融合

Vis: Park Sat: Golf
GT: Golf, Fuse: Golf



Vis: Park, Sat: Football Field
GT: Football, Fuse: Football



Vis: Themepark, Sat: Mall
GT: Mall, Fuse: Mall



Vis: Football, Sat: Tennis Court
GT: Tennis, Fuse: Tennis



Vis: Park, Sat: Residential
GT: Residential, Fuse: Residential



Vis: Beach, Sat: Themepark
GT: Themepark, Fuse: Themepark



Jiebo Luo, Jie Yu, Dhiraj Joshi and Wei Hao: Event Recognition: Viewing the World with a Third Eye, ACM Multimedia, pp.1071-1080, (2008).

Keita Yaegashi and Keiji Yanai: Can Geotags Help Image Recognition ?, (2009),

ジオタグ画像の位置情報を画像認識に利用.

Contributions of our work

- 特定のカテゴリの対象に関する世界各地の代表画像を自動的に検出
 - Eg.) “noodle” : Chinese noodle in Asia, spaghetti in Europe
 - 既存研究のように ランドマークに限定しない。カテゴリの代表画像なので、SIFTの対応点の数ではなく、BoFベクトルをクラスタリングして代表抽出
- 方法は新しくないが、目的が新しい。
 - 大規模ジオタグ画像DBの新しい利用法。
 - 「言語」「画像」の関係に「位置」を加えた新しい研究。

方法: 3つのステップ

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ *Generate image feature vectors*
- ✓ *Visual clustering*
- ✓ *Select most relevant clusters*

2) 代表的地域を検出

- ✓ *Clustering based on geographic locations*

3) 各代表地域それぞれの代表画像セットを生成

- ✓ *Generate the PLSA topic vectors*
- ✓ *Aggregate photos according to the distribution of mixture topics*

方法： 第一ステップ (1)

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ **Generate image feature vectors**

- ✓ Visual clustering

- ✓ Select most relevant clusters

Bag-of-visual-words
as
image representation

2) 代表的地域を検出

- ✓ Clustering based on geographic locations

3) 各代表地域それぞれの代表画像セットを生成

- ✓ Generate the PLSA topic vectors

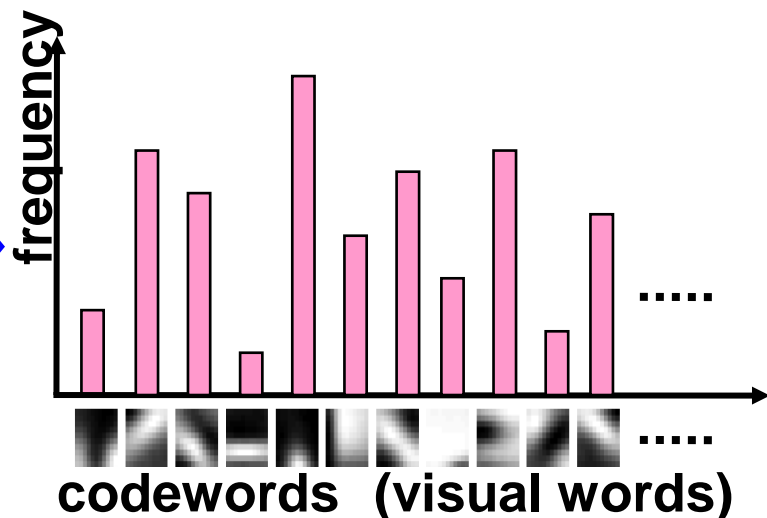
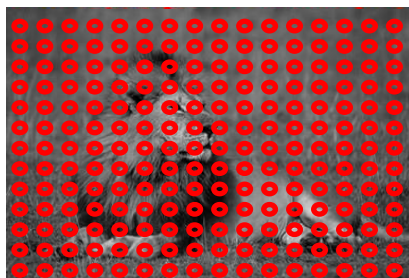
- ✓ Aggregate photos according to the distribution of mixture topics

[画像の表現法]

Bag-of-visual-words (BoVW)

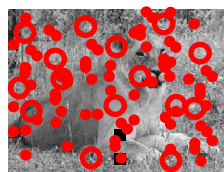
■ 画像を visual word の出現頻度ヒストグラムで表現

1. 各画像について、数千個の特徴点を抽出。
2. SIFT記述子により特徴点周辺パターンをSIFT特徴ベクトル(128次元)として抽出
3. 予め求められた visual words (codewords)に基づいてSIFT特徴ベクトルをベクトル量子化
4. 画像毎にヒストグラムを作成



Visual words の求め方

- 学習画像(正例, 負例)を用意し, SIFT特徴ベクトルを全画像から抽出 (枚数が多い場合は, ランダムサンプリング)
- k-means クラスタリングを実行



各クラスタの中心が “visual words”

v_i v_i

v_i v_i

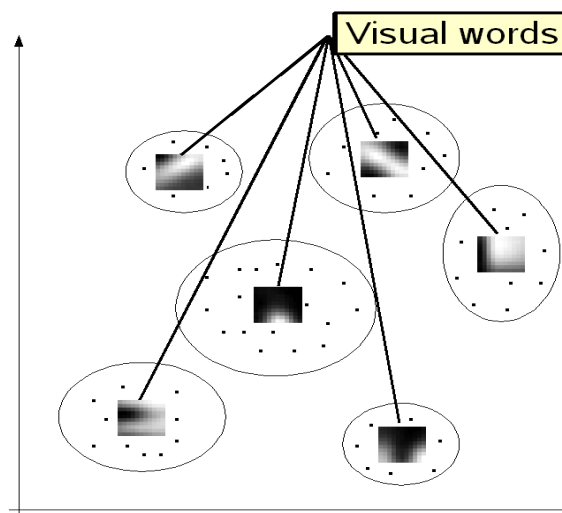
v_i v_i

v_i v_i

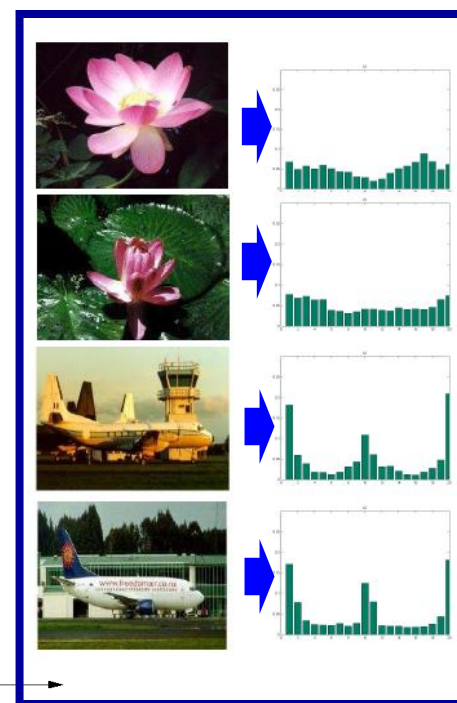
v_i v_i

v_i v_i

SIFT vectors



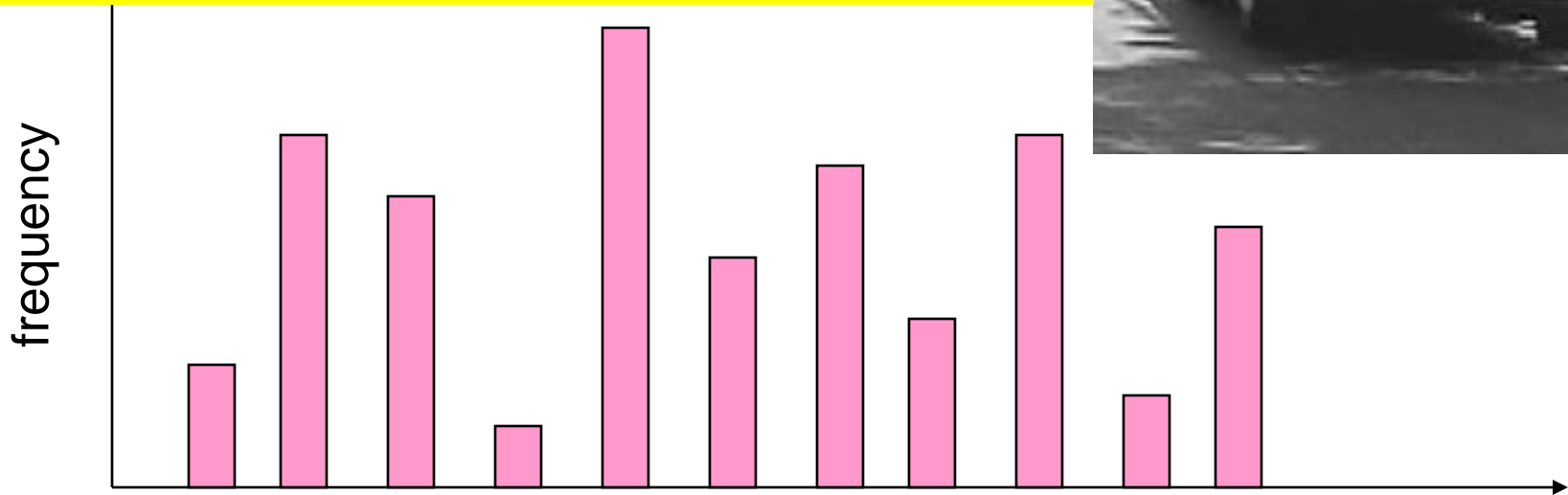
特徴空間



“Visual words” は, 代表的な局所パターンに相当する.

Bag-of-visual-words 表現

コードパターンの出現頻度 (ヒストグラム)
によって画像を表現。
テキスト解析のBag-of-wordsのアナロジー



【参考文献】

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.

方法： 第一ステップ (2)

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ Generate image feature vectors
- ✓ **Visual clustering**
- ✓ Select most relevant clusters

k-means clustering over bag-of-visual-words vectors.

2) 代表的地域を検出

- ✓ Clustering based on geographic information

3) 各代表地域それぞれの代表画像を生成

- ✓ Generate the PLSA topic vectors
- ✓ Aggregate photos according to the distribution of mixture topics

In the experiments, we set k to 200 for 2000 images.

方法： 第一ステップ (3)

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ Generate image feature vector
- ✓ Visual clustering
- ✓ **Select most relevant clusters**

Evaluate the intra-cluster similarity.

2) 代表的地域を検出

- ✓ Clustering based on geograph

Discard the clusters with small members ($m < 10$) or small similarity

3) 各代表地域それぞれの代表画像セットを生成

- ✓ Generate the PLSA topic vectors
- ✓ Aggregate photos according to the distribution of mixture topics

Select relevant clusters based on the intra-cluster similarity

- The intra-cluster similarity is the average similarity between the images that belong to the same cluster.

$$SIM(\mathbb{C}) = \frac{\sum_{P_i, P_j \in \mathbb{C}, i \neq j} \underline{sim(P_i, P_j)}}{nC_2}$$

$sim(P_i, P_j)$ is the similarity between two images P_i and P_j computed by cosine metric.

$$sim(P_i, P_j) = \frac{V_i \cdot V_j}{\sqrt{|V_i| |V_j|}}$$

- Discard clusters with small members ($m < 10$)
- Select the top 40 clusters in terms of $SIM(\mathbb{C})$ after sorting $SIM(\mathbb{C})$ in the descending order.

Noise filtering by selecting 40 clusters out of 200 clusters for 2000 images

方法： 第二ステップ。

1) Flickrから“X”画像を収集

- ✓ Generate image feature vectors
- ✓ Visual clustering
- ✓ Select most relevant clusters

K-means clustering over location vectors :
(longitude, latitude)
(k=5)

2) 代表的地域を検出

- ✓ Clustering based on geographic locations

3) 各代表地域それぞれの代表画像セットを生成

- ✓ Generate the PLSA topic vectors
- ✓ Aggregate photos according to the distribution of mixture topics

方法： 第三ステップ (1)

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ Generate image feature vectors
- ✓ Visual clustering
- ✓ Select most relevant clusters

2) 代表的地域を検出

- ✓ Clustering based on geographic locations

3) 各代表地域それぞれの代表画像セットを生成

- ✓ **Generate the PLSA topic vectors**
- ✓ Aggregate photos according to the distribution of mixture topics

Probabilistic Latent Semantic Analysis (PLSA)

- **テキスト解析向けの確率トピックモデル**
 - **Bag-of-words**表現された文書を確率的にトピック分類する. ソフトクラスタリング. GMMに類似.
 - トピック数は, 事前に指定する
 - トピックを z , 文書(画像)を d とすると, 各文書について $P(z|d)$ が求まる
- **Probabilistic Latent Semantic Analysis**
 - $P(w, d) = P(d) \sum p(w | z) P(z | d)$ をEMで推定

$$P(\text{word}, \text{image}) = P(\text{image}) \sum_{\text{topic}} p(\text{word} | \text{topic}) P(\text{topic} | \text{image})$$

確率的トピック

Mountain 10 topics



画像を
潜在トピック
の混合分布
で表現

$P(\text{topic}|\text{image})$

clusters: 10 word: mountain JUMP

0.15



0.24



0.03



0.05



0.10



0.01



0.01



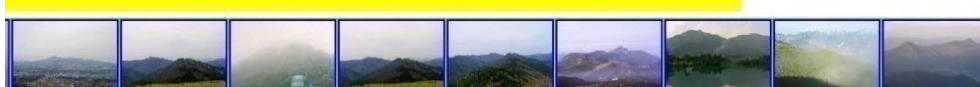
0.01



0.05



0.30



方法： 第三ステップ (2)

1) Flickrから“X”画像を収集し、ノイズ除去

- ✓ Generate image feature vectors
- ✓ Visual clustering
- ✓ Select most relevant clusters

2) 代表的地域を検出

- ✓ Clustering based on geograph

3) 各代表地域それぞれの

- ✓ Generate the PLSA topic vectors

- ✓ **Aggregate photos according to the distribution of mixture topics**

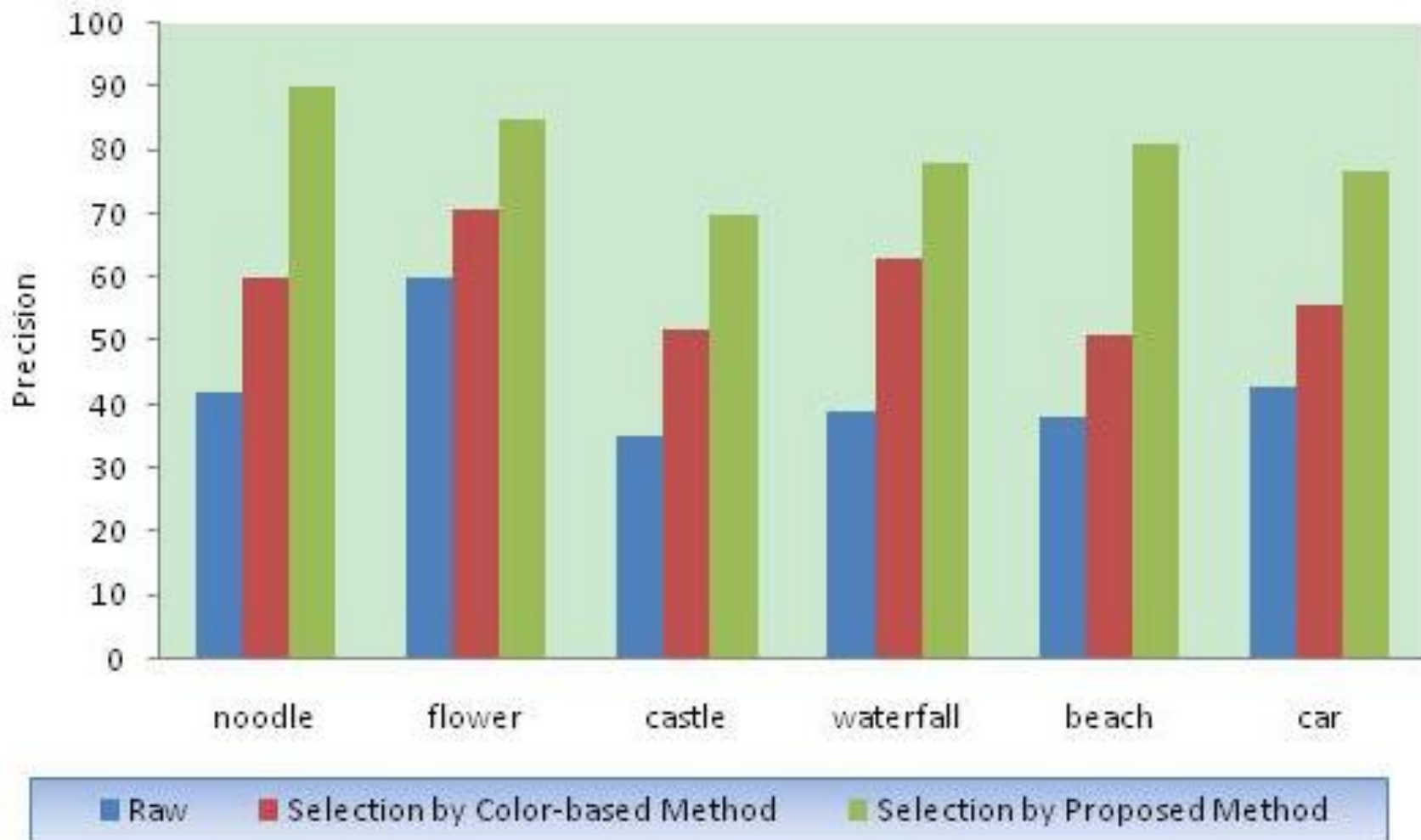
K-means clustering over topic vectors ($k=5$) for each region .
Select the largest cluster as a final result.

実験結果

- **“noodle”, “flower”, “castle”, “waterfall”, “beach”, “car”, “wedding cake” (7種類)**
 - ✓ For each concept, collect about 2000 geo-tagged photos from Flickr distributed evenly in the world wide areas
- **第一ステップ(ノイズ除去)の評価**
 - ✓ Evaluation on our proposed method for extracting the most relevant photos
 - ✓ Precision and Recall
 - ✓ Color-histogram-based method for comparison
- **あとは、結果を見てください。**

第一ステップ(ノイズ除去)の評価

■ 適合率(選ばれた画像のうち正しい画像の割合)



第一ステップ(ノイズ除去)の評価

■ *raw vs. color-based vs. BoVW*

Concepts	Raw photos From Flickr	Selection by Color-based Method	Selection by Proposed Method
noodle	2080 (42)	769 (60, 54)	752 (90, 80)
flower	2225 (60)	703 (71, 37)	705 (85, 45)
castle	1848 (35)	780 (52, 61)	761 (70, 81)
waterfall	1901 (39)	689 (63, 59)	672 (78, 70)
beach	1917 (38)	824 (51, 58)	813 (80, 90)
TOTAL/AVG.	9971 (43)	3765 (59, 54)	3703 (81, 73)

Precision

Recall

[Example of results] "noodle"

[Please choose your options] Keyword: noodle District: [v] [Jump] [Animation Control] [Auto Play] [Stop] [Continue]

Representative photos for "noodle"

Map data ©2008 AND, Geocentre Consulting, NFGC, ZENRIN

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

[Please choose your options] Keyword: noodle District: [v] [Jump] [Animation Control] [Auto Play] [Stop] [Continue]

Representative photos for "noodle"

Map data ©2008 AND, Geocentre Consulting, NFGC, ZENRIN

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

[Please choose your options] Keyword: noodle District: [v] [Jump] [Animation Control] [Auto Play] [Stop] [Continue]

Representative photos for "noodle"

Map data ©2008 AND, Geocentre Consulting, NFGC, ZENRIN

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

[Please choose your options] Keyword: noodle District: [v] [Jump] [Animation Control] [Auto Play] [Stop] [Continue]

Representative photos for "noodle"

Map data ©2008 AND, Geocentre Consulting, NFGC, ZENRIN

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

[Please choose your options] Keyword: noodle District: [v] [Jump] [Animation Control] [Auto Play] [Stop] [Continue]

Representative photos for "noodle"

Map data ©2008 AND, Geocentre Consulting, NFGC, ZENRIN

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

[Example of results] "noodle"

[Please choose your options] Keyword: District: [Animation Control]

Map Satellite Hybrid Terrain

Representative photos for "noodle"

The screenshot shows a web application interface for searching for 'noodle' across Europe. At the top, there is a search bar with the keyword 'noodle' and a 'District' dropdown set to '3'. To the right are 'Animation Control' buttons: 'Auto Play', 'Stop', and 'Continue'. Below the search bar is a map of Europe with several blue location pins. A callout box is open over a location in Italy, displaying a photo of spaghetti and the coordinates [41.900659, 12.514436]. To the right of the map is a grid titled 'Representative photos for "noodle"', which contains 40 small images of various noodle dishes. At the bottom of the interface, there is a copyright notice: '© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan'.

Taiwanese spaghetti photos in the Europe area
Japanese ramen photos in the Japan area

[Example of results] "flower"



[Example of results] “flower”

[Please choose your options] Keyword: District: [Animation Control]

Map Satellite Hybrid Terrain

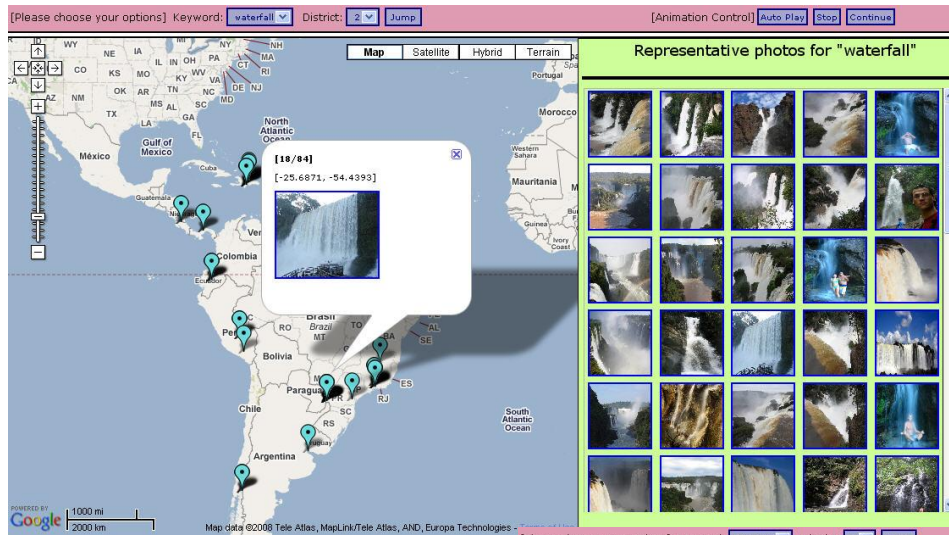
[11/68]
[52.269316, 4.546076]

Representative photos for "flower"

© 2008 Yanai Lab., Department of Computer Science, The University of Electro-Communications, Tokyo, Japan

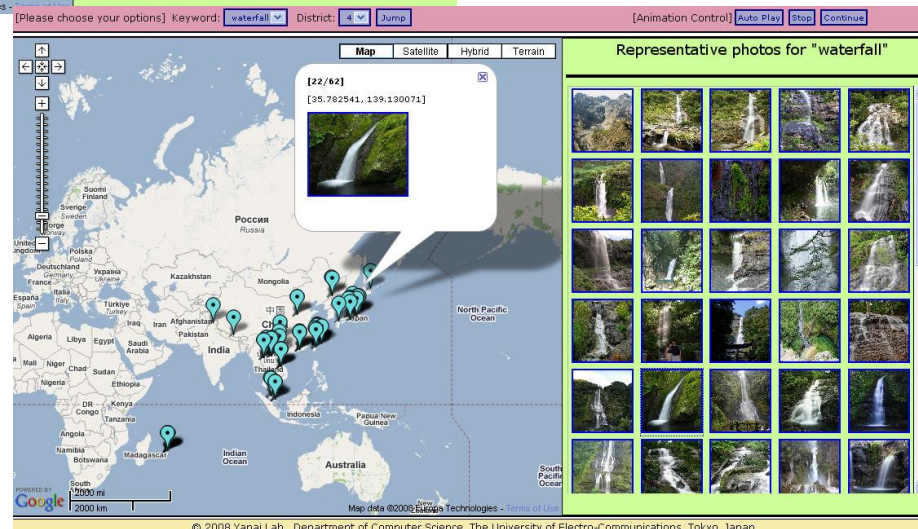
Netherlands national flower flower “Tulip”

[Example of results] “waterfall”



“Powerful” waterfalls
in South America

“Beautiful” waterfalls
in Asia



まとめ & 今後の課題

- 特定のカテゴリーに関する地域毎の代表的画像を抽出する方法を提案.
- 今後の課題
 - ✓ 大規模な実験
 - ✓ 地域を分類の階層化. **世界→アジア→日本→東京**
 - ✓ 画像の「代表度」を定量化し, ランキング.
 - ✓ 文化の違いの自動検出. 単語による差異の大小分析
 - ✓ **【最終目標】 日本「ラーメン」文化マイニング**
(ジオタグ画像でラーメン日本地図を作って, 文化差異の分析)

位置と画像特徴のエントロピー

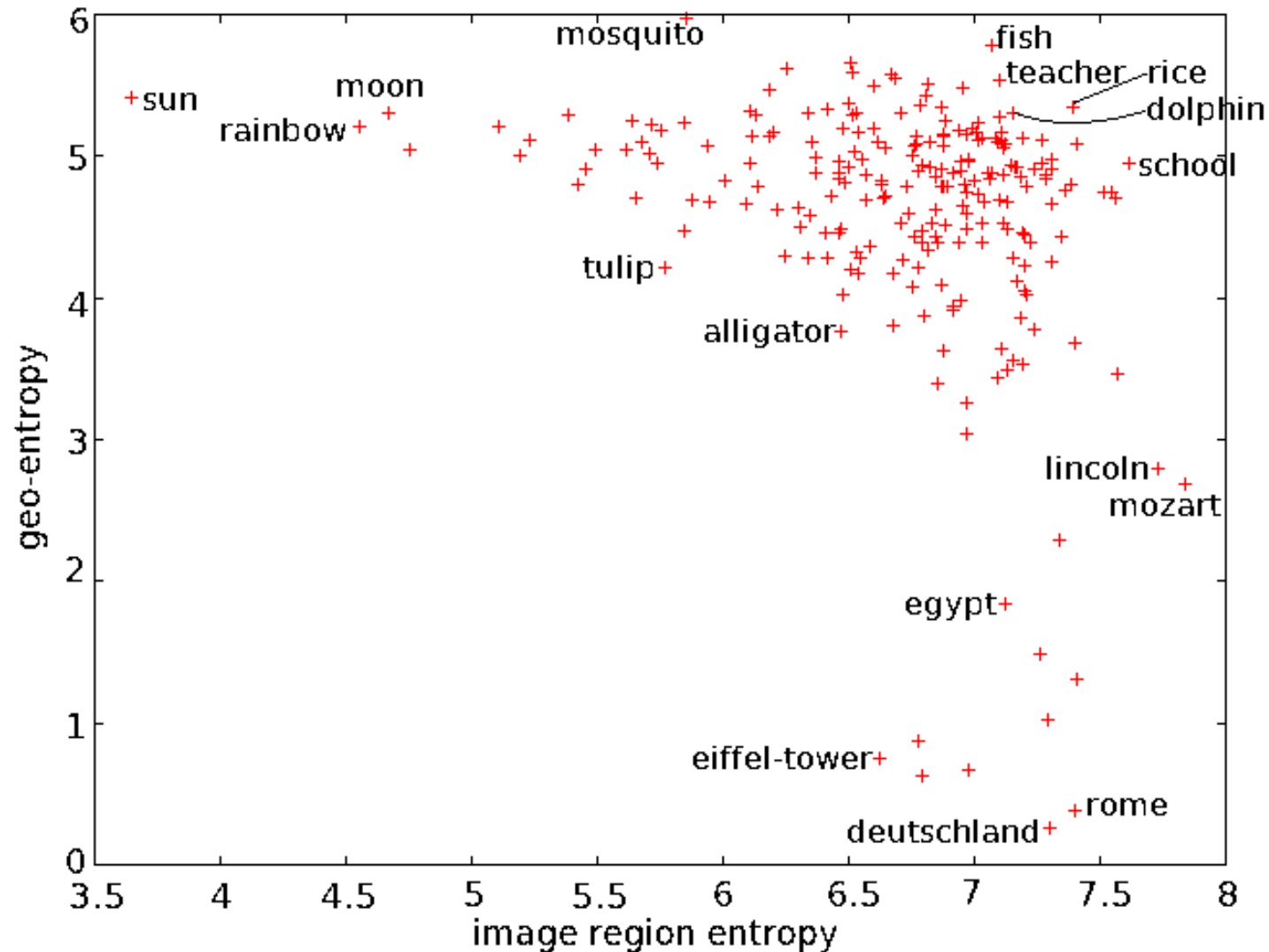




Fig. 5. "Mozart" images over the world.

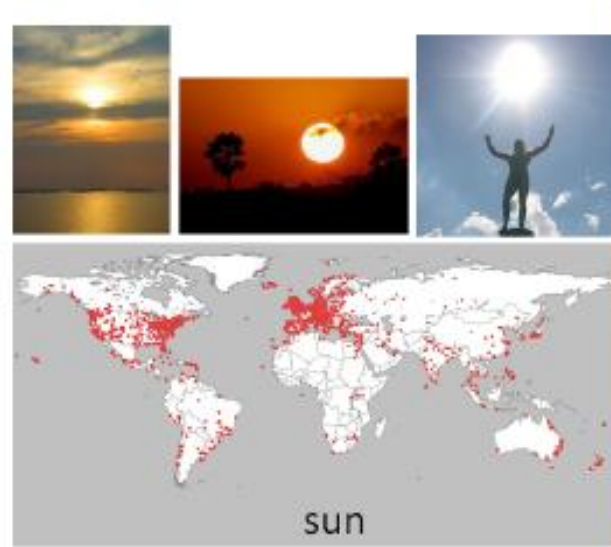


Fig. 6. "Sun" images over the world.

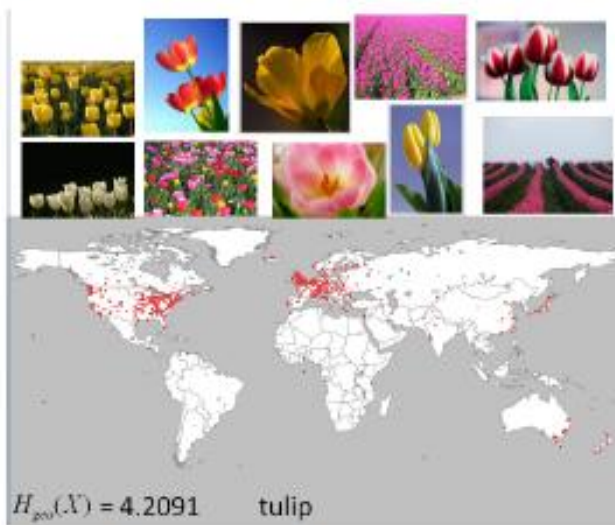


Fig. 7. "Tulip" images over the world.



Fig. 8. "Rice" images over the world.



For more results, please access:

<http://img.cs.uec.ac.jp/yanai/ASRP/>

Thank you !