

写真属性と画像特徴を用いたホット撮影スポット・ アノテーション

Hot Photo-Spot Annotation using photo attributes and image features

小関 基徳¹ 熊野 雅仁^{2*} 亀井 貴行¹ 小野 景子² 木村 昌弘²
Motonori Koseki¹ Masahito Kumano² Takayuki Kamei¹
Keiko Ono² Masahiro Kimura²

¹ 龍谷大学大学院理工学研究科電子情報学専攻

¹ Division of Electronics and Informatics, Graduate School of Science and Technology, Ryukoku University

² 龍谷大学 理工学部 電子情報学科

² Department of Electronics and Informatics, Faculty of Science and Technology, Ryukoku University

Abstract: In our previous work, we presented a method of extracting a pair of a major photo-spot and its hot-period, which is called a *hot photo-spot*, from a large number of geotagged photographs with timestamps that many people have taken. However, as for explaining the hot photo-spots extracted, it was in general difficult to annotate them clearly since each of them can have a variety of photos. In this paper, we propose a method of explaining each hot photo-spot by classifying the photos in it based on their image features and attributes such as their geotags and timestamps. Using real data from “Flickr data”, we experimentally demonstrate the effectiveness of the proposed method.

1 はじめに

近年、写真を撮影する際、どこで撮影したか(地理情報)を、写真に付与できるデジカメやカメラ付き携帯端末が一般化し始めている。また、Flickr¹など、多くの写真共有サイトが賑わいを見せており、写真と共に地理情報を登録できる機能を備えているため、Web空間に共有化された、膨大な地理情報付き写真データが蓄積され続けている。写真は、撮影者の心をつかむ対象に遭遇したとき撮影されることが多いことから、写真が単なる記録ではなく、撮影者の何らかの意見を内在化させていると考えることができる。つまり、大量の写真群は、意見の集合と見なすことができ、写真そのものから得られる情報や、付随する情報をうまく集約すれば、集合知が得られる可能性がある[味八木 10]。

一方、Web空間に電子化された観光情報が溢れるに従い、計算機科学の領域では「観光」が注目されており[川村 10, 松原 11]、近年、観光や旅行支援への応用も期待できる新しいアプローチとして、地理情報付き写真群を用いる研究が脚光を浴びている。Crandallらは、

大量の地理情報付き写真と、写真の画像特徴を用いて、空間的なクラスタリングを行い、多くの人が訪れる人気スポットや、ランドマークのある主要地域が得られることを示した[Crandall 09]。この地理情報付き写真群を用いる空間に着目した研究は、魅力的な地域を抽出し[Kisilevich 10b]、視覚的に探索する研究[Kisilevich 10a]や、写真に付与された文書情報も利用することで、地域ごとの地理的トピックを抽出し、地域間の文化を比較して新たな知識を発見する研究[Yin 11]や、観光マップを生成する研究[王 11]などに派生している。

また、Crandallらは、同一の撮影者が同日に複数の写真を撮影した場合、写真が撮影された時間情報を追跡し、地理情報と併用することで、撮影地点の軌跡が得られることも示した。この地理情報および時間情報を用いる研究は、旅行する人々の写真撮影行動から旅行行動をマイニングする研究[Arase 10]や、旅行の計画を支援する研究[Yin 10]、旅行計画の経路を生成する研究[Lu 10]などに派生している。

Crandallらが抽出した人気スポットは、実空間に局在する地域に写真群が密集することを重点に置くため、年間を通じて人々が訪れ写真を撮る地域が優先的に抽出される傾向がある。しかし、写真の撮影スポットを推薦する問題を考える場合、「どこで」という地理情

*連絡先： 龍谷大学
滋賀県大津市瀬田大江町横谷 1-5
E-mail:{kumano,kono,kimura}@rins.ryukoku.ac.jp

¹<http://www.flickr.com/>

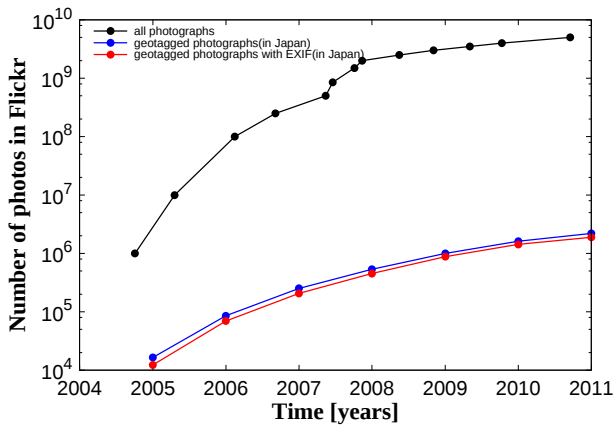


図 1: Flickr の登録写真数変遷

報だけでなく、「いつ」という時期の情報が欠けていると、旬のある撮影期間を逃しかねない。我々は、多数の撮影者が生み出した地理情報および時間情報付き大規模写真データから、撮影者の意見が反映され、他の地域と比較して普段とは逸脱して顕著に撮影数が増える、集合知的観点を背景とした格別な地域と期間のペアをホット撮影スポットと呼んで、その自動抽出問題に取り組み、ホット撮影スポットが観光スポットとして推薦できる可能性を示した [熊野 12]。

ところで、ホット撮影スポットは、空間的クラスタリングを行う上で、mean-shift 法 [Comaniciu 02] を用いているが、そのパラメータとして、カーネル関数のカーネル幅 h をあらかじめ与える必要があり、ホット撮影スポットには、 h の設定により、多数の写真が含まれることがある。ホット撮影スポット内に多数の写真が含まれる場合、一つ一つの写真に Geo-tag 情報が付随しているため、より詳細に調べれば、サブ撮影スポットが存在する可能性がある。しかし、ホット撮影スポットを観光スポットとして推薦することを考えた場合、その地域と期間に含まれる写真群が混沌とした未整理状態であるとわかりにくいという問題がある。

例えば、撮影地点がほぼ同じでも、撮影対象が異なる場合、複数の撮影対象が混在する。また、イベントなどにおいても、撮影地点はほぼ同じであるが、時間帯により異なる対象が興味の対象になる場合が考えられる。また、熊野らは、旬のある時期について、日を基準とした期間に関する結果を示したが、日の出は朝の時間帯、夜景などでは夜の時間帯が旬となる場合が考えられ、時間帯の違いによる撮影対象が混在する。つまり、撮影場所や時間帯、撮影対象の違いによる混沌とした状態を整頓することで、ホット撮影スポット内のサブ撮影スポットが明瞭になることが期待される。

Naaman らは、ひとりの撮影者による地理および時間情報付き写真データを地域およびイベントの観点から

整理する研究を行っている [Naaman 04]。しかし、ホット撮影スポットでは、複数の異なる嗜好を持つ撮影者による多様な写真が混在しているため、ホット撮影スポットを明瞭に説明することが困難であった。本研究では、ホット撮影スポットの写真群について、写真の撮影位置、撮影時間と、画像特徴に基づいてホット撮影スポットを明瞭に分類することで、ホット撮影スポットを説明するアノテーション付与法を提案する。

そこで、まず 2 章で写真共有サイトについて説明し、3 章でホット撮影スポット抽出法について述べる。また、4 章で提案法を述べ、5 章で Flickr から収集した日本全土を含む写真の実データを用いた実験と考察を行い、提案法の有効性を示す。そして、6 章でまとめる。

2 写真共有サイトと写真属性

写真共有サイトは、ユーザどうしがコミュニケーションを取る意味で、ソーシャルネットワーキングサイト (SNS) の一つであるが、主にユーザが撮影した写真を実世界の事象を捉えた情報源と見なせば、ユーザが生み出した写真を通じて情報を発信することから、ソーシャルメディアであるとも言える。2012 年現在、世界中に、数多くの写真共有サイトが存在するため、Web 空間には膨大な写真が蓄積され続けている。一方、デジタル写真には、EXIF (Exchangeable image file format) 情報が埋め込まれており、撮影時刻などの基本的な情報の他に、GPS 情報や撮影条件などの多数の情報が含まれている。Flickr は、主要写真共有サイトの一つであるが、写真の EXIF 情報が自動登録されたり、ユーザが付与したタグ情報など、多数の属性情報も共有化して閲覧できる機能を持つ。

図 1 は、Flickr に登録された各年ごとの写真総数である。図 1 の黒線は、登録総数の変遷であり、青線は、日本列島が含まれる地域の位置を指定して得られる写真総数である。ここで、Flickr から得られる位置に、少なくとも二つの観点による位置情報が含まれることを説明しておく。一つ目は GPS に基づいた撮影位置である。デジタル写真に埋め込まれたメタ情報に GPS に基づく撮影位置情報がある場合、Flickr では位置情報の登録拒否を設定していないかぎり、写真登録時にジオタグとして自動記録され、変更できなくなる。二つ目は、Flickr ユーザが地図ツール上の位置を任意に指定して登録した位置であり、登録された写真に一度も Geo-tag が記録されていない場合に設定が可能となる。二つ目の場合でも撮影位置を登録することはできるが、ユーザに一任されるため、全く関係のない位置を登録することもできる。つまり、Flickr の写真に付随する位置情報は、必ずしも撮影位置を示さないこともあるが、EXIF 情報に記録される GPS 情報は、撮影時点で記録される

ため、情報を改ざんしないかぎり、撮影位置を示すと言える。図 1 の赤線は、日本列島が含まれる地域で、EXIF 情報を持つ写真総数の変遷である。本研究では、写真の属性情報として、撮影位置と撮影時間を用いる。

3 ホット撮影スポット抽出

3.1 ホット撮影スポットとホットスポット写真

正の整数 T に対して、 T 日の期間 $[1, T]$ 内に撮影された写真データ全体の集合を、

$$\mathcal{D}_T = \{d_n; n = 1, \dots, N\}$$

とする。ここに、各写真データ d_n には、Geo-tag 情報 x_n 、時間情報 t_n が付随しており、そのことを明記するために、

$$d_n = (x_n, t_n), \quad (n = 1, \dots, N)$$

と記述する。ただし、 $x_n = (x_{n,1}, x_{n,2})$ であり、 $x_{n,1}$ と $x_{n,2}$ はそれぞれ写真 d_n が撮影された緯度と経度、 t_n は d_n が撮影された日、 N は写真データの総数である。

緯度と経度の情報を用いれば、地球表面上の点は 2 次元 Euclid 空間 $\mathbf{R}^2$ 内の領域

$$\Omega = [-\pi/2, \pi/2] \times [-\pi, \pi] \subset \mathbf{R}^2$$

上の点と同一視される。写真データ集合 $\mathcal{D}_T$ から、多くの写真が撮影される人気撮影スポットが近接して存在する地域 $R_m (\subset \Omega)$, ($m = 1, \dots, M$) を抽出し、その地域において格別の期間 $I_m = [T_{m,0}, T_{m,1}]$, ($m = 1, \dots, M$)、すなわち、他の地域と比較して顕著に人々がその地域で写真を撮影している期間を検出する。各 R_m を主要撮影地域、 I_m を R_m のホット撮影期間と呼ぶ。ここに、 M は抽出した主要撮影地域の総数であり、 R_m は半径 h_0 のある円板近傍に含まれる Ω 内の領域、 $1 \leq T_{m,0} < T_{m,1} \leq T$, ($m = 1, \dots, M$) である。ただし、 $h_0 (> 0)$ は、主要撮影地域のサイズを規定するパラメータである。ここで、 R_m と I_m のペア (R_m, I_m) をホット撮影スポットと呼び、与えられた T 日間の写真データ集合 $\mathcal{D}_T$ からホット撮影スポット群 $\{(R_m, I_m); m = 1, \dots, M\}$ を抽出する。

また、($m = 1, \dots, M$) に対し、地域 R_m 内で期間 I_m に撮影された写真群、つまり、 D_m に属する写真をホット撮影スポット m のホットスポット写真と呼ぶ。

$$D_m = \{d_n = (x_n, t_n) \in \mathcal{D}_T; x_n \in R_m, t_n \in I_m\},$$

3.2 ホット撮影スポットの数理モデル

次に、ホット撮影スポットの数理モデルを示す。写真データ集合 $\mathcal{D}_T$ から主要撮影地域 R_m , ($m = 1, \dots, M$)

を抽出する手法において、人々が写真をどの場所で撮影するのかに関する確率分布に対して、その確率密度関数を極大にする点の近傍が主要撮影地域であるとモデル化する。ただし、極大値が比較的小さいものについては、主要撮影地域とは考えないことにする。そのような確率密度関数の推定に対して、ノンパラメトリックアプローチであるカーネル密度推定

$$\hat{p}(x) = \frac{1}{Nh^2} \sum_{n=1}^N G(\| (x - x_n) / h \|^2), \quad (x \in \mathbf{R}^2) \quad (1)$$

を考える。ここに、 $\|\cdot\|$ は $\mathbf{R}^2$ の Euclid ノルム、 $G(s)$ はカーネル関数であり、Epanechnikov カーネルや Gaussian カーネルなどを利用する。また、 $h (> 0)$ は、主要撮影地域のサイズを規定するパラメータとして、対象とする問題のスケール（解像度）に応じてユーザが事前に指定するものとする。

我々は、Crandall らの研究 [Crandall 09] に従い、 $\mathcal{D}_T$ に属する各写真の撮影場所 x_n , ($n = 1, \dots, N$) を初期値としてミーンシフト法を適用し、式 (1) の確率密度関数 $\hat{p}(x)$ の極大値を与える点を推定するとともに、 $\mathcal{D}_T$ に属する写真のクラスタリングを行う。 $\hat{p}(x)$ の極大値を与える点として推定されたもの全体を $\{\hat{c}_m; m = 1, \dots, M'\}$ とし、各 m に対して $\hat{c}_m$ に収束した x_n , ($n = 1, \dots, N$) の全体を、

$$X_m = \{x_{n(m,j)}; j = 1, \dots, N_m\}, \quad (m = 1, \dots, M')$$

とする。ただし、 $|X_1| \geq \dots \geq |X_{M'}|$ とする。ここで、 $|X_m| \geq \mu_0$ を満たす $m \in \{1, \dots, M'\}$ の最大値 M を求める。ここに、 μ_0 はユーザが指定するパラメータである。次に、各 $m \in \{1, \dots, M\}$ に対して、 $\hat{c}_m$ を中心とし X_m を含む最小の円板近傍と領域 Ω との共通部分 R_m を求める。そして、 $\{R_1, \dots, R_M\}$ を主要撮影地域として出力する。抽出された各主要撮影地域 R_m に対して、そのホット撮影期間 $I_m = [T_{m,0}, T_{m,1}]$ を検出する手法を提案する。ここに、 $T_{m,0}$ と $T_{m,1}$ は $T_{m,0} < T_{m,1}$ なる T 以下の自然数である。

任意の $m \in \{1, \dots, M\}$ に対して、 $q_m(t)$ を R_m 内で第 t 日に撮影された写真の数とし、各 $q_m(t)$ が

$$q_k(t) = q_k^*(t) + q_0(t) \quad (2)$$

のように分解されるとモデル化する。ここに、 $q_0(t)$ は m に依存しない正整数で、地域によらず一般的に第 t 日に撮影される写真数を表す確率変数である。また、 $q_m^*(t)$ は、地域 R_m に特徴的な撮影動向を表すもので、通常の日には m によって異なる正定数値 $w_{m,0}$ をとり、ホット撮影期間 I_m において $w_{m,0}$ より大きい正定数値をとる階段関数である。ただし、各 R_m に対して、ホット撮影期間 I_m は複数個（例えば、 $I_{m,1}, I_{m,2}, \dots$ ）存在し得るが、それらの任意の 2 つの交わりは空集合である。ま

た、 $m \neq m'$ ならば、 R_m と $R_{m'}$ のホット撮影期間は一致しないとする。

任意の主要撮影地域に対して、そのホット撮影期間の候補全体は $\mathcal{J} = \{J = [T_0, T_1]; T_0, T_1 \in \mathbb{Z}, 1 \leq T_0 < T_1 \leq T\}$ であり、それらを、

$$\mathcal{J} = \{J_i; i = 1, \dots, T(T-1)/2\}$$

と番号づけする。ここで、各 R_m におけるホット撮影期間（すなわち、他の地域と比較して顕著に多数の写真が撮影された期間）を効率的に検出するために、撮影された写真の数に関して、地域 R_m , ($m = 1, \dots, M$) と期間 J_i , ($i = 1, \dots, T(T-1)/2$) の独立性を検定する。具体的には、まず Fisher 直接確率検定に従って、 R_m と独立性が低い（すなわち、Fisher 直接確率の値が小さい）期間を候補 $\mathcal{J}$ から探索する。ところで、 R_m に対する Fisher 直接確率の値が小さい期間は、他の地域と比較して顕著に少数の写真が撮影された期間という場合もあり得るので、Fisher 直接確率検定で検出された期間に対して、さらにその期間で撮影された写真数をも考慮し、 R_m におけるホット撮影期間を抽出する。

3.3 ホット撮影期間の抽出

まず、Fisher 直接確率検定に従って、地域 R_m , ($m = 1, \dots, M$) と期間 J_i , ($i = 1, \dots, T(T-1)/2$) の独立性を検定する。表 1 のような R_m と J_i に関する 2×2 分割表を考えよう。ここで、 N は写真の総数、 p_m は領域 R_m に属する写真の数、 p'_i は期間 J_i に含まれる写真の数、 $p_{m,i}$ は R_m に属し J_i に含まれる写真の数、 $p_{\bar{m},i}$ は R_k に属し J_i に含まれない写真の数、 $p_{\bar{m},\bar{i}}$ は R_k に属さず J_i に含まれる写真の数、 $p_{\bar{m},\bar{i}}$ は R_k に属さず J_i に含まれない写真の数を、それぞれ表す。このとき、

$$p_{k,i} + p_{\bar{m},i} = p_m, \quad p_{\bar{m},i} + p_{\bar{k},\bar{i}} = N - p_m,$$

$$p_{m,i} + p_{\bar{m},i} = p'_i, \quad p_{k,\bar{i}} + p_{\bar{k},\bar{i}} = N - p'_i$$

である。Fisher 直接確率検定では、Fisher 直接確率

$$F_{m,i} = \sum_{j=p_{m,i}}^{\min(p_m, p'_i)} \frac{\binom{p_m}{j} \binom{N-p_m}{p'_i-j}}{\binom{N}{p'_i}} \quad (3)$$

表 1: 2×2 分割表

	J_i	$\bar{J}_i$	
R_k	$m_{k,i}$	$m_{\bar{k},i}$	m_k
$\bar{R}_k$	$m_{\bar{k},i}$	$m_{\bar{k},\bar{i}}$	$N - m_k$
	m'_i	$N - m'_i$	N

が大きいほど、 R_m と J_i の独立性が高いと検定される。我々は、各 R_m に対して $p_{m,i} \geq \phi_m$ なる J_i を、Fisher 直接確率 $F_{m,i}$ の小さい順に「 $I_{m,1}, I_{m,2}, \dots$ 」とランキングし、「 $I_{m,1}$ を R_m の第 1 ホット撮影期間、 $I_{m,2}$ を R_m の第 2 ホット撮影期間、 $\dots$ 」として抽出する。ここに、 $\phi_m (> 0)$ はユーザが指定するパラメータである。

Fisher 直接確率 $F_{m,i}$, ($m = 1, \dots, M; i = 1, \dots, T(T-1)/2$) は、原理的には式 (3) に従ってナイーブに直接計算することにより求めることが可能だが、 N と T が大きくなると膨大な計算量が必要になると考えられる。そこで我々は、

$$f(\ell, j) = \log \left(\frac{\ell}{j} \right), \quad (\ell = 1, \dots, N; j = 0, 1, \dots, \ell)$$

を、漸化式

$$f(\ell, j) = \begin{cases} 0 & (j = 0) \\ f(\ell, j-1) + \log(\ell - j + 1) - \log(j) & (j \geq 1) \end{cases} \quad (4)$$

を用いて計算し、これらから Fisher 直接確率 $F_{k,i}$ を、

$$F_{m,i} = \sum_{j=p_{m,i}}^{\min(p_m, p'_i)} \exp(f(p_m, j) + f(N - p_m, p'_i - j) - f(N, p'_i)) \quad (5)$$

と計算することを提案する。式 (5) では指数値を計算する前に対数値 $f(\ell, j)$ の加算および減算を実行しているので、提案法が計算精度の劣化を抑制することも期待できる。以下に、ホット撮影スポットの抽出アルゴリズムを示す。

Algorithm 提案法

- 1: p_m を求める;
- 2: 式 (4) から $\{f(p_m, i); 1 \leq i \leq p_m\}, \{f(N - p_m, i); 1 \leq i \leq N - p_m\}, \{f(N, i); 1 \leq i \leq N\}$ を求める;
- 3: $i := 1$;
- 4: **while**($i \leq T(T-1)/2$) **do** /*期間 J_i の $F_{m,i}$ を計算*/
- 5: R_m と J_i に関する 2×2 分割表を構築 (表 1 を参照);
- 6: **if** $p_{m,i} < \phi_m$ **then**
- 7: **goto** step10;
- 8: **end if**
- 9: 式 (5) から $F_{m,i}$ を求める;
- 10: $i := i + 1$;
- 11: **end while**
- 12: $F_{m,i}$ の小さい順に J_i をランキングし、 R_m におけるホット撮影期間 $I_{m,1}, I_{m,2}, \dots$ を出力;

さらに我々は、Fisher 直接確率 $F_{m,i}$ に基づいて主要撮影地域とホット撮影期間候補のペア (R_m, J_i) をランキングすることにより、与えられた T 日間の写真データ集合 $\mathcal{D}_T$ から、格別なホット撮影スポット群を抽出する。

4 提案法

4.1 概要

本研究では、写真群を明瞭に分類する上で、各写真に付随する Geo-tag、画像データ、撮影時間を用いて、1. Geo-tag 情報、2. 画像特徴情報、3. 時間差情報の三つの情報に着目し、ホット撮影スポット m ごとに、未整理のホットスポット写真群 D_m を三情報の組み合わせによる類似度に基づいてクラスタリングすることで、ホット撮影スポットを説明するアノテーションを付与する手法を提案する。

4.2 アノテーション

Geo-tag による実距離の近さのみによって写真群をクラスタリングし、色分けラベルを付けて地図上に配置すると、サブ撮影スポットの存在が明瞭になる。また、画像特徴の近さを用いてクラスタリングし、色分けラベルを付け、Geo-tag 情報を用いて地図上に配置すると、異なる場所に同じような写真が撮影可能なサブ撮影スポットの存在を示すことができる。さらに、撮影時間の近さを用いれば、朝の撮影スポット、夜の撮影スポットなどが異なるクラスとして分類できる。

一方、Geo-tag 情報と画像特徴情報の両方の近さを複合してラベル付けをすれば、ほぼ同じ撮影位置でも、複数のクラスに分割される可能性があり、撮影対象の違いによる分類が可能になる場合が考えられる。さらに、Geo-tag 情報、画像特徴情報、撮影時間差情報の三つを統合すれば、同じ場所でも、異なる対象、異なる時間帯のサブ撮影スポットであるというラベル付けが可能になると思われる。

つまり、撮影位置の近さ、写真に映る画像の近さ、撮影時間の近さの指標の組み合わせに基づいて、分類を行った写真群が、うまく撮影スポットの特徴を表しているならば、クラスに付与されたラベルが、写真群を説明することから、本研究では、そのラベルをアノテーションと見なす。以下に、三つの情報による類似度の算出法と組み合わせ法を示す。

4.2.1 Geo-tag 情報に基づく類似度

ホット撮影スポット m 内の写真群 D_m について、各写真間の実距離に基づく類似度 S_G が $0 \sim 1$ 内になるよう正規化を行う。そこで、 $dist_G(u, v)$ ただし $(u \neq v)$ を d_u と d_v との Geo-tag に基づく距離とし、その最大値を $dist_G^{max}$ 、最小値を $dist_G^{min}$ としたとき、 S_G を次のように定義する。

$$S_G(u, v) = 1 - (dist_G(u, v) - dist_G^{min}) / (dist_G^{max} - dist_G^{min})$$

4.2.2 画像特徴に基づく類似度

近年、画像の特徴を表現する手法として、Visual words (keypoints) が注目されている [Csurka 04]。本研究では、計算速度の速い SURF (Speeded-Up Robust Features) 特徴量 [Bay 08] をホット撮影スポット内の写真群 D_m から抽出し、 K -means クラスタリングを適用し、得られたセントロイドを Visual Word とすることで、個々の写真を K 次元の Visual words ヒストグラムで表現する。ここで、 $sim_I(u, v)$ ただし $(u \neq v)$ を d_u と d_v との $\cos$ 類似度とし、その最大値を sim_I^{max} 、最小値を sim_I^{min} としたとき、画像の類似度 S_I を次のように定義する。

$$S_I(u, v) = (sim_I(u, v) - sim_I^{min}) / (sim_I^{max} - sim_I^{min})$$

4.2.3 撮影時間差に基づく類似度

異なる二枚の写真 u と v ただし $(u \neq v)$ が撮影された時間を PT_u 、 PT_v とするとき、撮影時間の近さを表すため、24 時間を円周に割り当て、 PT_u と PT_v の時間差を、円周の時計回りと、反時計回りでそれぞれ求め、差が小さい方の絶対値を撮影時間差 $diff_T(u, v)$ とする。ここで、その最大値を $diff_T^{max}$ 、最小値を $diff_T^{min}$ としたとき、画像の類似度 S_T を次のように定義する。

$$S_T(u, v) = 1 - (diff_T(u, v) - diff_T^{min}) / (diff_T^{max} - diff_T^{min})$$

4.2.4 類似度の組み合わせ

本研究では、三つの類似度を組み合わせた合成類似度 S' を以下の組み合わせで統合した。

1. $S'_G = S_G / w$
2. $S'_I = S_I / w$
3. $S'_T = S_T / w$
4. $S'_{G,I} = S_G / w + S_I / w$
5. $S'_{G,T} = S_G / w + S_T / w$
6. $S'_{I,T} = S_I / w + S_T / w$
7. $S'_{G,I,T} = S_G / w + S_I / w + S_T / w$

ただし、類似度が一つの場合 $w = 1$ 、二つ統合する場合 $w = 2$ 、三つ統合する場合 $w = 3$ とする。

4.3 クラスタリング

本研究では、 D_m からクラスを抽出するクラスタリング手法として、クラス数を自動決定できる Newman クラスタリング [Clauset 04] を用いる。この手法は、大規模な複雑ネットワークのグラフに内在するコミュニティ構造を高速に抽出する手法であるため、グラフ表現されたネットワークを構築する必要がある。そこ

で、 D_m 内の写真をノードとし、異なる二つの写真の類似度 S' を重みとするリンクを張り、完全グラフを構築して、閾値 S_h 以下のリンク削除する。得られたグラフ構造のうち、最大連結成分 G をクラスタリングの対象として、サブ撮影スポットを抽出する。

Newman クラスタリングでは、グラフ内に潜在的に存在するコミュニティ i 内のリンク密度が高く、コミュニティ間のリンク密度が低い状態を良いクラスタ構造と見なす。クラスタリングの精度を測るモジュール性指標 (modularity) $Q > 0$ を導入し、 Q が 0 になるランダムネットワークに対し、コミュニティ構造の存在を反映した Q を最大にするクラスタ数を自動決定できる。

$$Q = \sum_i (e_{ii} - a_i^2)$$

ここで、 e_{ii} は「コミュニティ i 内のノードと i 内の別のノード間のリンク総和」の「総リンク数」に対する割合を表し、 a_i は「コミュニティ i 内のノードから出ているリンク総和」の「総リンク数」に対する割合である。また、 $\Delta Q_{ij} = 2(e_{ij} - a_i a_j)$ を計算することで、高速化も行われている。本研究では、重み付きネットワーク G に Newman クラスタリング [Zhou 10] を適用する。

5 実験

写真共有サイト Flickr から収集した大量の地理情報および時間情報の付随する写真データを用いて、ホット撮影スポットの抽出を行い、ホット撮影スポットごとにクラスタの抽出実験を行った。

5.1 実験データ

日本国内のホット撮影スポットでアノテーション付与を行うため、写真共有サイト Flickr から、日本列島が含まれる矩形領域 (緯度:25.8 ~ 45.8, 経度:126.2 ~ 146.8) に含まれる 2010 年 1 月 1 日から 2010 年 12 月 31 日までの 1 年間の撮影位置・撮影時間付き写真データを収集した。ただし、日本国内に焦点を当てるため、矩形領域に入り込む他国の写真データを除いた。その結果、548,922 枚の写真データが得られた。また、本研究では、空間クラスタリングにおいて、Yin[Yin 11] らに従い、最小領域で一人の撮影者が何度撮影しても 1 度として数えた。本研究では、最小領域を 1 辺 10m の矩形領域とした。その結果、写真数 162,933 枚のデータセットとなった。

実験の前段階として、(ホット撮影スポット) の項で説明した手順に従って h を 100m として空間クラスタリングを行った。その結果、 $K'=24,954$ 箇所の主要撮影地域が得られた。 $\mu_0=100$ としたところ、 $K=205$ となっ

た。続いて、fisher 検定で各地域のホット撮影期間を検出してホット撮影スポットを得た。全ホット撮影スポットの写真数 $|D_T|$ は 10,185 枚となった。本研究では、この D_T を実験用のデータセットとする。

5.2 クラスタリングの比較

Visual Word の語数を決める $K=1000$ として Visual Words ヒストグラムを写真 d_n ごとに算出した。また、三つの類似度を用いて組み合わせた七つの類似度 S' をそれぞれ用いて、七種類の完全グラフを構築し、 $S' < S_h$ となるリンクを切断し、それぞれ最大連結成分をクラスタリングの対象とした。本研究では、リンクを切断する閾値 $S_h=0.8$ として実験を行った。

5.2.1 評価法

まず、本研究では、三つの情報による $S'_{G,I,T}$ を用いたクラスタリングによるクラスタへのラベル付けを提案法と定め、クラスタ C_k , ($k = 1, \dots, N_k$) を正解データとする。ここで、三つの情報を用いることの効果を評価するため、他の S' を用いた場合と相違があるか否かを調べる。尚、 S' の異なるクラスタリングでは、それぞれ得られるクラスタ数が異なることが予想されるが、本研究では、Micro average precision を用いることで、クラスタ数の違いによる問題を解消する。一方、他の組み合わせの特徴量を使ったクラスタリングの結果を比較対象として C'_l , ($l = 1, \dots, L_l$) とする。このとき、比較するデータのクラスタ C'_l に対して交わり具合が最大となる C_k との適合率を C'_l の得点 $f(l)$ とする。

$$f(l) = \max |C'_l \cap C_k|$$

そして、次式で比較するデータの正解データに対する Micro average precision P_{ma} を算出する。

$$P_{ma} = \sum_{n=1}^L \frac{f(l)}{N}$$

5.2.2 クラスタリング結果の評価

$S'_{G,I,T}$ に対する他の S' についての P_{ma} について、205 のホット撮影スポットで算出した値の平均値を表 2 に示す。表 2 より、最もクラスタの傾向が近いのは「画像特徴」+「時間情報」を組み合わせた $S_{I,T}$ であるが、それでも 0.62 程度の値であり、他の S' を含め、得られたクラスタ群はいずれとも異なっている独自性があることがわかる。つまり、三つを組み合わせた提案法は、部分的に共通する情報を用いていても、他の S' を用いた場合と異なるクラスタを形成していると言える。

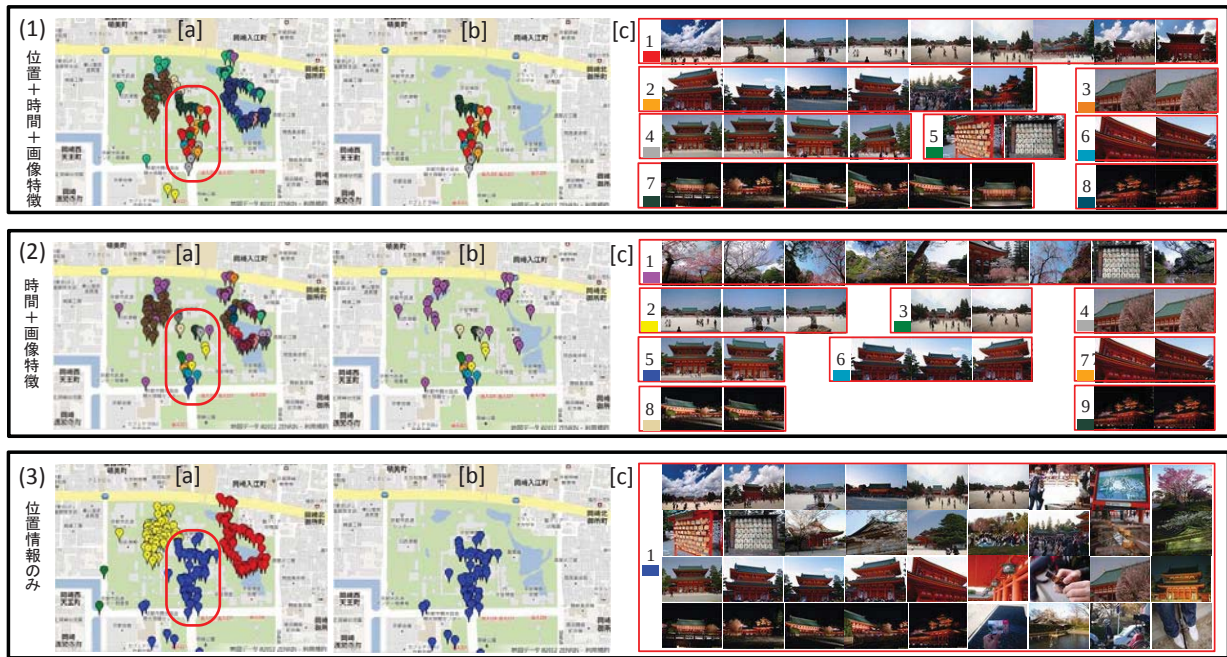


図 2: 提案法 (1) と比較法 (2)(3) (ホット撮影スポット：京都、平安神宮)

表 2: micro average precision の平均値ランキング

	画像 + 位置 + 時間情報に対する値
画像 + 時間情報	0.628234731
位置情報のみ	0.538383589
位置 + 時間情報	0.537968227
画像 + 位置情報	0.526942926
時間情報のみ	0.441309438
画像情報のみ	0.409779561

5.3 実験結果

図 2(1) は、三つの情報を用いてクラスタリングした提案法の結果を抜粋した一例である。図 2(1)[a] は、クラスタリングの結果、ホット撮影スポットの一つである、京都、平安神宮周辺のホットスポット写真群を Geo-tag を用いて Google Map に配置したものである。説明上、図 2(1)[a] の赤枠の地域に着目する。図 2(1)[b] は、その赤枠の中に全部もしくは一部が含まれる 8 クラスタであり、図 2(1)[c] は、その 8 クラスタの写真を抜粋したものである。平安神宮は、日中に撮影されることも多いものの、ライトアップがあり、夜景も数多く撮影される観光スポットである。位置情報により、クラスタは近い位置関係のものが抽出されているが、ほぼ同じ撮影地域でも、画像特徴の効果から異なる撮影対象が別のクラスタに分かれ、写真の印象が近いものが同じクラスタに含まれる様子がわかる。また、ほぼ同じ撮影地域でも、昼と夜の写真はよく分離されており、同

じ印象の写真が同クラスタに含まれることから、提案法は、うまく分類され、何が撮影できるかが明瞭に理解しやすいクラスタを形成していることがわかる。

一方、図 2(2) は、提案法に最も近かった「撮影時間」と「画像特徴」の近さを用いたクラスタリングの結果である。図 2(2)[a] では、図 2(1)[a] とほぼ同じ地域に赤枠を設定し、その赤枠の中に全部もしくは一部が含まれる 9 クラスタを示したものが図 2(2)[b] である。図 2(2)[b] 中、紫色のクラスタが、赤枠内に一部存在しているが、紫色のクラスタは、平安神宮のあちこちに分散している。これは「位置」の近さが反映されていないことが原因であると思われる。この紫色のクラスタは、図 2(2)[c] の 1 の赤枠に対応しており、同クラスタ内に含まれる写真は、位置が異なっている、同じような印象の写真が集まっていることがわかる。これは、紫色クラスタのような写真を好む者からすれば、平安神宮にいくつも同じような写真が撮影できるスポットがあることを示していると考えられることも可能である。しかし、図 2(2)[c] の (2) や (3) は、位置の情報がないためか、実際のところ極めて近い撮影位置であるにもかかわらず、画像特徴がうまく働かなかったためか、過分割された例である。この点においては、提案法では、図 2(1)[c] の 1 クラスタに対応しているが、位置情報の効果からか、過分割が行われず、一つのクラスタとなつて同様の印象の写真が整理されている印象がある。

また、図 2(3) は、提案法に対し、2 番目に近かった位置情報のみを使った比較法である。これまでと同様に、図 2(3)[a] で、他とほぼ同じ地域を抽出した結果が

図 2(3)[b] であるが、これは図 2(3)[c] の 1 クラスタのみが抽出された。このクラスタは、時間帯の違い、撮影対象の違いに関係なく、多様な写真が混在して、どのようなクラスタなのか説明しにくい印象を受ける。以上から、提案法の有効性が伺える。

6 まとめ

本研究では、写真データに付随する Geo-tag 情報と撮影時間情報、さらに画像特徴に着目して三つの情報を統合することで、ホット撮影スポットをうまく分類し、提案法の有効性を示した。今後は、より効果的な整理法を探索するとともに、観光スポットを推薦する上で、デジタル写真に付随する数多くの EXIF 情報などを用いて、より有効な整頓法を探索する予定である。

参考文献

- [Arase 10] Arase, Y., Xie, X., Hara, T., and Nishio, S.: Mining People's Trips from Large Scale Geo-tagged Photos, in *Proceedings of the 18th International Conference on Multimedia*, pp. 133–142 (2010)
- [Bay 08] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L.: Speeded-Up Robust Features (SURF), *Comput. Vis. Image Underst.*, Vol. 110, No. 3, pp. 346–359 (2008)
- [Clauset 04] Clauset, A., Newman, M. E. J., and Moore, C.: Finding community structure in very large networks, *Physical Review E*, pp. 1–6 (2004)
- [Comaniciu 02] Comaniciu, D. and Meer, P.: Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603–619 (2002)
- [Crandall 09] Crandall, D. J., Backstrom, L., Huttenlocher, D., and Kleinberg, J.: Mapping the world's photos, in *Proceedings of the 18th International Conference on World Wide Web*, pp. 761–770 (2009)
- [Csurka 04] Csurka, G., Dance, C. R., Fan, L., Willamowski, J., and Bray, C.: Visual categorization with bags of keypoints, in *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22 (2004)
- [Kisilevich 10a] Kisilevich, S., Mansmann, F., Bak, P., Keim, D. A., and Tchaikin, A.: Where Would You Go on Your Next Vacation? - A Framework for Visual Exploration of Attractive Places, in *GeoProcessing 2010*, pp. 21–26 (2010)
- [Kisilevich 10b] Kisilevich, S., Mansmann, F., and Keim, D. A.: P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos, in *1st International Conference on Computing for Geospatial Research & Application* (2010)
- [Lu 10] Lu, X., Wang, C., Yang, J.-M., Pang, Y., and Zhang, L.: Photo2Trip: generating travel routes from geo-tagged photos for trip planning, in *Proceedings of the 18th International Conference on Multimedia*, pp. 143–152 (2010)
- [Naaman 04] Naaman, M., Song, Y. J., Paepcke, A., and Garcia-Molina, H.: Automatic Organization for Digital Photographs with Geographic Coordinates, in *Proceedings of ACM/IEEE-CS JCDL Joint Conference on Digital Libraries*, pp. 53–62 (2004)
- [Yin 10] Yin, H., Lu, X., Wang, C., Yu, N., and Zhang, L.: Photo2Trip: an interactive trip planning system based on geo-tagged photos, in *Proceedings of the 18th International Conference on Multimedia*, pp. 1579–1582 (2010)
- [Yin 11] Yin, Z., Cao, L., Han, J., Zhai, C., and Huang, T.: Geographical Topic Discovery and Comparison, in *Proceedings of the 20th International Conference on World Wide Web*, pp. 247–256 (2011)
- [Zhou 10] Zhou, T. C., Ma, H., Lyu, M. R., and King, I.: UserRec: A User Recommendation Framework in Social Tagging Systems, in Fox, M. and Poole, D. eds., *AAAI*, AAAI Press (2010)
- [王 11] 王 佳な, 野田 雅文, 高橋 友和, 出口 大輔, 井手 一郎, 村瀬 洋: Web 上の大量の写真に対する画像分類による観光マップの作成, *情報処理学会論文誌*, Vol. 52, No. 12, pp. 3588–3592 (2011)
- [熊野 12] 熊野 雅仁, 小関 基徳, 小野 景子, 木村 昌弘: 地理および時間情報をもつ写真データに基づいたホット撮影スポットの抽出, *情報処理学会論文誌*, Vol. 5, No. 3, pp. 41–53 (2012)
- [松原 11] 松原 仁: 特集: 「観光と知能情報」にあたって, *人工知能学会誌*, Vol. 26, No. 3, p. 225 (2011)
- [川村 10] 川村 秀憲, 鈴木 恵二, 山本 雅人, 松原 仁: 観光情報学, *情報処理*, Vol. 51, No. 6, pp. 642–648 (2010)
- [味八木 10] 味八木 崇, 暦本 純一: 集合知センシングによる実世界インタフェース, *情報処理*, Vol. 51, No. 7, pp. 775–781 (2010)