

## 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

熊野 雅仁<sup>†1</sup> 小関 基徳<sup>†1</sup>  
小野 景子<sup>†1</sup> 木村 昌弘<sup>†1</sup>

人々を惹きつける実世界の撮影スポットを発見する問題は、観光産業などへの応用が期待できるため、重要である。しかし、撮影スポットは、空間的に局在する地域を特定するだけでは、旬のある撮影期間を逃す可能性がある。本研究では、数多くの人々が撮影した、撮影場所を示す地理情報や撮影時間の情報が付与された大量の写真データを用いて、実世界の空間に局在する主要撮影地域と、時間軸上に局在し、他の地域と比較して顕著にその地域で写真が撮影されている格別な期間とのペアをホット撮影スポットと呼んで、その局在する実時空間の抽出問題に取り組む。そこで、本研究ではホット撮影スポットの数理モデルと、その効率的な抽出法を提案する。そして、Flickrから収集した実データを用いた実験により、提案法の有効性を示す。

### Extracting Hot Photo-spots from Geotagged Photographs with Timestamps

MASAHITO KUMANO,<sup>†1</sup> MOTONORI KOSEKI,<sup>†1</sup> KEIKO ONO<sup>†1</sup>  
and MASAHIRO KIMURA<sup>†1</sup>

From the point of view of sightseeing industry, discovering good photo-spots in the real world is an important problem. While a popular photo-spot in general means a spatially localized region, it should have its own *hot-period*, a period of time during which it can provide more interesting photographs than other photo-spots. In this paper, we address the problem 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. We propose a mathematical model for hot photo-spots, and present a method of efficiently identifying them. Using real 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、Flickr<sup>\*1</sup>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。この写真共有サイトのユーザは、実世界の対象や出来事を記録する撮影者であるだけでなく、Web空間に情報発信することからソーシャルメディア<sup>1)</sup>と見なすことができ、さらに、実世界をセンシングすることからソーシャルセンサとして捉えることもできる<sup>2)</sup>。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<sup>3),4)</sup>、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。

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<sup>5)</sup>。Crandallら<sup>6)</sup>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、多くの人々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。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、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<sup>7)</sup>や、観光への応用として、観光マップを生成する研究<sup>8)</sup>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<sup>9)</sup>や、旅行の計画を支援する研究<sup>10)</sup>、旅行計画の経路を生成する研究<sup>11)</sup>などに派生している。しかし、以上の経路の概念は、写真に付随する時間情報が利用されているものの、主に空間上の概念と言える。つまり、これまでの研究は、いずれも空間的な観点が重点的に扱われていると言え、まだ時期に関する観点が十分に検討されていない。

<sup>†1</sup> 龍谷大学 理工学部 電子情報学科

Department of Electronics and Informatics, Ryukoku University

\*1 <http://www.flickr.com/>

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。例えば京都の醍醐時は、世界遺産の一つであり、四季折々の見どころがあるもの、他の名所でない特徴として、豊臣秀吉が贅をつくし、各地の有名な桜を醍醐寺に集め、一箇所で見るができるため、4月に最も撮影者を引き寄せる。また、愛知の豊田市小原は、奥地でありながら、11月頃に撮影者を引き寄せる。それは、四季桜という4月と10月末の年に2回咲くめずらしい桜が植えられているため、10月末から11月中旬に紅葉と桜を同時に見るのできる格別の場所だからである。ある場所のある時期に他の地域と比べて普段以上に人が集まり、多数の写真が撮影されるとき、地理および時間情報付きの写真群は、撮影者の意見を代弁し、旬のある名所であることを教えてくれる。つまり、撮影スポットに着目する場合、時期の存在は欠かせない。時期に着目した先駆的な研究としては、文書時系列が与えられたとき、トピックと期間に着目して $\chi^2$ 検定を用いることで存在期間が付随するホットトピックを抽出する研究<sup>12)</sup>や、並行して存在する複数の異なるトピックの時間的遷移を空間的に可視化する研究<sup>13)</sup>などがある。しかし、写真データを対象として、地理情報と時間(期間)のペアを扱い、実世界に局在する時空間を抽出する研究は、これまでなかった。

本研究では、時間軸上に局在する期間を抽出する研究の流れと、実空間内に局在する地域を抽出する研究の流れを拡張し、空間内の地理情報と時間軸上の期間のペアに着目して、実世界に局在する局在実時空間を抽出するという問題を論じる。ここで、本研究では、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局在実時空間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む。そして、ホット撮影スポットの数理モデルと、構築された数理モデルに基づいた、その効率の良い抽出法を提案する。そこで、2節で問題の定式化を与え、3節で提案法を詳説する。さらに、4節で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、5節でまとめる。

## 2. 問題の定式化

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、地理情報  $x_n$ 、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。Flickr など多くの写真共有サイトから、このような写真データが大量に収集可能であることに注意しておく。

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

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

上の点と同一視される。我々は、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  $R_k (\subset \Omega)$ , ( $k = 1, \dots, K$ ) を抽出し、その地域において格別の期間  $I_k = [T_{k,0}, T_{k,1}]$ , ( $k = 1, \dots, K$ )、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ことを考える。各  $R_k$  を主要撮影地域、 $I_k$  を  $R_k$  のホット撮影期間と呼ぶ。ここに、 $K$  は抽出した主要撮影地域の総数であり、 $R_k$  は半径  $h_0$  のある円板に含まれる  $\Omega$  内の領域、 $1 \leq T_{k,0} < T_{k,1} \leq T$ , ( $k = 1, \dots, K$ ) 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。我々は、 $R_k$  と  $I_k$  のペア  $(R_k, I_k)$  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  $\{(R_k, I_k); k = 1, \dots, K\}$  を抽出する問題を考える。

地域  $R_k$  内で期間  $I_k$  に撮影された写真群を、

$$D_k = \{d_n = (x_n, t_n) \in \mathcal{D}_T; x_n \in R_k, t_n \in I_k\}, \quad (k = 1, \dots, K)$$

とする。 $D_k$  に属する写真を  $R_k$  のホットスポット写真と呼ぶ。

## 3. 提案法

ホット撮影スポットの数理モデルを構築し、与えられた  $T$  日間の写真データ集合  $\mathcal{D}_T = \{d_n = (x_n, t_n); n = 1, \dots, N\}$  から、ホット撮影スポット群  $\{(R_k, I_k); k = 1, \dots, K\}$  を抽出する手法を提案する。本節では、まず、主要撮影地域  $R_k$ , ( $k = 1, \dots, K$ ) の数理モデルとそれらを抽出する手法を与え、次に、ホット撮影期間  $I_k$ , ( $k = 1, \dots, K$ ) の数理モデルとそれらを効率的に抽出する手法を提案する。

### 3.1 主要撮影地域の抽出法

写真データ集合  $\mathcal{D}_T$  から主要撮影地域  $R_k$ , ( $k = 1, \dots, K$ ) を抽出する手法を与える。まず我々は、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<sup>6)</sup> 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  $\{\hat{c}_k; k = 1, \dots, K'\}$  とし、各  $k$  に対して  $\hat{c}_k$  に収束した  $x_n$ , ( $n = 1, \dots, N$ ) の全体を、

$$X_k = \{x_{n(k,j)}; j = 1, \dots, N_k\}, \quad (k = 1, \dots, K')$$

とする。ただし、 $|X_1| \geq \dots \geq |X_{K'}|$  とする。我々は、 $|X_k| \geq \mu_0$  を満たす  $k \in \{1, \dots, K'\}$  の最大値  $K$  を求める。ここに、 $\mu_0$  はユーザが指定するパラメータである。次に、各  $k \in \{1, \dots, K\}$  に対して、 $\hat{c}_k$  を中心とし  $X_k$  を含む最小の円板と領域  $\Omega$  との共通部分  $R_k$  を求める。そして、 $\{R_1, \dots, R_K\}$  を主要撮影地域として出力する。

### 3.2 ホット撮影期間の検出法

抽出された各主要撮影地域  $R_k$  に対して、そのホット撮影期間  $I_k = [T_{k,0}, T_{k,1}]$  を検出する手法を提案する。ここに、 $T_{k,0}$  と  $T_{k,1}$  は  $T_{k,0} < T_{k,1}$  なる  $T$  以下の自然数である。

任意の  $k \in \{1, \dots, K\}$  に対して、 $q_k(t)$  を  $R_k$  内で第  $t$  日に撮影された写真の数とする。我々は、各  $q_k(t)$  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  $k$  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、 $q_k^*(t)$  は、地域  $R_k$  に特徴的な撮影動向を表すもので、通常の日  $k$  によって異なる正定数値  $w_{k,0}$  をとり、ホット撮影期間  $I_k$  において  $w_{k,0}$  より大きい正定数値をとる階段関数である。ただし、各  $R_k$  に対して、ホット撮影期間  $I_k$  は複数個（例えば、 $I_{k,1}, I_{k,2}, \dots$ ）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、 $k \neq k'$  ならば、 $R_k$  と  $R_{k'}$  のホット撮影期間は一致しないとす。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

表 1  $2 \times 2$  分割表  
Table 1  $2 \times 2$  contingency table.

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

$\mathcal{J} = \{J = [T_0, T_1]; T_0, T_1 \in \mathbf{Z}, 1 \leq T_0 < T_1 \leq T\}$  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。我々は、各  $R_k$  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  $R_k$ , ( $k = 1, \dots, K$ ) と期間  $J_i$ , ( $i = 1, \dots, T(T-1)/2$ ) の独立性を検定することを考える。具体的には、まず Fisher 直接確率検定に従って、 $R_k$  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、 $R_k$  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、 $R_k$  におけるホット撮影期間を抽出する。以下に、 $R_k$  のホット撮影期間を検出するための提案手法を詳述する。

まず、Fisher 直接確率検定に従って、地域  $R_k$ , ( $k = 1, \dots, K$ ) 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  $R_k$  と  $J_i$  に関する  $2 \times 2$  分割表を考えよう。ここで、 $N$  は写真の総数、 $m_k$  は領域  $R_k$  に属する写真の数、 $m'_i$  は期間  $J_i$  に含まれる写真の数、 $m_{k,i}$  は  $R_k$  に属し  $J_i$  に含まれる写真の数、 $m_{k,\bar{i}}$  は  $R_k$  に属し  $J_i$  に含まれない写真の数、 $m'_{k,i}$  は  $R_k$  に属さず  $J_i$  に含まれる写真の数、 $m'_{k,\bar{i}}$  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、

$$m_{k,i} + m_{k,\bar{i}} = m_k, \quad m'_{k,i} + m'_{k,\bar{i}} = N - m_k, \quad m_{k,i} + m'_{k,i} = m'_i, \quad m_{k,\bar{i}} + m'_{k,\bar{i}} = N - m'_i$$

である。Fisher 直接確率検定では、Fisher 直接確率

$$F_{k,i} = \sum_{j=m_{k,i}}^{\min(m_k, m'_i)} \frac{\binom{m_k}{j} \binom{N - m_k}{m'_i - j}}{\binom{N}{m'_i}} \quad (3)$$

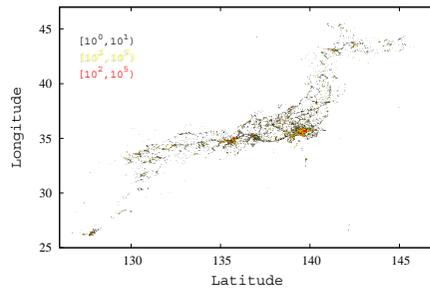


図1 Flickr データセットにおける撮影位置分布。  
Fig. 1 Geographic locations of photographs in the Flickr dataset.

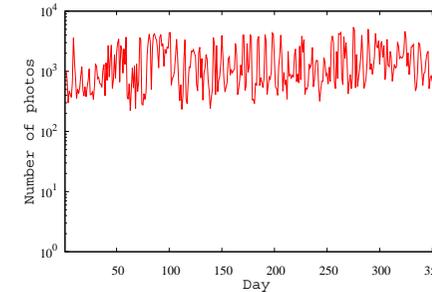


図2 Flickr データセットにおける 1 日ごとの写真数の変遷。  
Fig. 2 Daily fluctuation of the number of photographs in the Flickr dataset.

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

Fisher 直接確率  $F_{k,i}$ , ( $k = 1, \dots, K; i = 1, \dots, T(T-1)/2$ ) 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、

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

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。

さらに我々は、Fisher 直接確率  $F_{k,i}$  に基づいて主要撮影地域とホット撮影期間候補のペア  $(R_k, J_i)$  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. 実データによる実験評価

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付き写真を用いて、提案法の有効性に関する実験を行った。

##### 4.1 実験データ

日本国内で提案法の性能を検証するため、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。図 1 は、写真データ群の写真撮影位置の分布を点で示したものであり、撮影地点の色は、写真数 0 枚を白として、写真の数を三つの色クラス  $\{ [10^0, 10^1):$  黒、 $[10^1, 10^2):$  黄、 $[10^2, 10^5):$  赤  $\}$  に分けて示した。図 1 より、京都近郊や、特に東京近郊の都市部で数多く撮影されているが、撮影地点は国土の形状が見えるほど日本全土に分布していることがわかる。

また、図 2 は、写真データ群の撮影時間情報に基づいて、1 年間 1 日ごと (横軸) に撮影された写真数 (縦軸) を示したものである。図 2 より、1 年を通じて撮影されていることがわかる。本研究では、主要撮影地域抽出においては、Yin<sup>7)</sup> に従い、最小領域で一人の撮影

表 2 提案法により抽出されたホット撮影スポット (Top 10).  
Table 2 Hot photo-spots extracted by the proposed method (Top 10).

Rank	都市	施設・地域 (写真の主な内容)	代表撮影位置 (緯度・経度)	撮影期間 (日数)	写真数
1	京都	醍醐寺 (桜)	34.95154090384615, 135.8187998269231	4/2-4/5(4)	169
2	東京	北野丸公園 (桜)	35.69242769444443, 139.7495188888889	3/31-4/7(8)	223
3	神奈川	小田原城 (桜)	35.25080959756099, 139.15348648780497	4/8-4/8(1)	115
4	東京	昭和記念公園 (コスモス)	35.714454, 139.39490104	10/17-10/17(1)	131
5	千葉	京成バラ園 (バラ)	35.73144322500001, 140.08629146666667	11/3-11/3(1)	122
6	愛知	小原 (四季桜・紅葉)	35.25913421538461, 137.30022104615378	11/20-11/20(1)	147
7	山梨	富士吉田 (富士山)	35.37871445454546, 138.74417633333331	7/23-7/23(1)	91
8	京都	平安神宮 (夜桜)	35.01567406504066, 135.7824145691056	4/3-4/8(6)	176
9	東京	台場 (モーターカーレース)	35.622330133333335, 139.77391376666657	10/2-10/2(1)	126
10	京都	仁和寺 (桜)	35.02922719685039, 135.71334140157478	4/11-4/14(4)	132



図 4 提案法によるランキング 1 位のホット撮影スポットにおける写真例 (醍醐寺).  
Fig. 4 Example of the photographs in the hot photo-spot of rank 1 extracted by the proposed method (Daigoji temple).

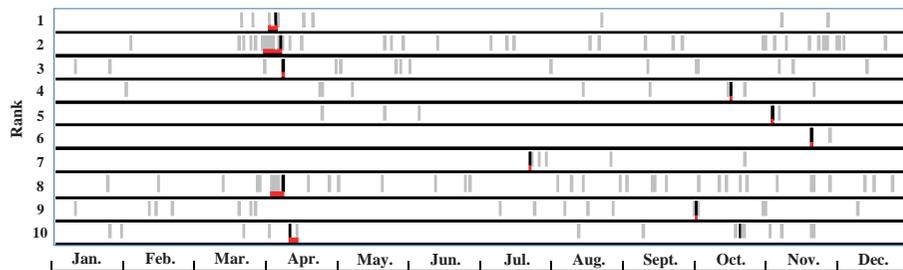


図 3 提案法によるホット撮影スポットのランキングトップ 10 における日ごとの写真数変動.  
Fig. 3 Daily fluctuations of the numbers of photographs in the top 10 hot photo-spots by the proposed method.

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、写真数は 162,933 枚となった。これを本稿では Flickr dataset と呼ぶ。

#### 4.2 ホット撮影スポットの抽出結果

提案法に基づくホット撮影スポットの抽出実験を行った。 $G(s)$  は Epanechnikov カーネルを用いて  $h$  を 100m とし、主要撮影地域抽出したところ、 $K'=24,954$  箇所の  $R_k$  が得られ、 $\phi_0=100$  としたところ、 $K=205$  となった。ホット撮影期間の抽出実験の結果、ホット撮影スポット候補ランキングのトップ 10 を示したものが表 2 である。表 2 には、各ホット撮影スポット候補の代表撮影位置 (緯度・経度) と撮影期間 (日数) を示した。ただし、表 2 の都市と地域・施設や写真の主な内容の情報は、代表撮影位置が含まれる施設やイベント等を手で調べて記載した。表 2 から、提案法では、東京だけでなく日本各地を捉えており、また、写真の主な内容から、目標となる限られた時期と相関がある対象として、開花時期があ

る桜やバラ、コスモスの咲く地域と期間が抽出されている。また、車の魅力を伝えるイベントとして毎年開催される、車好きの心を惹きつけるトップカテゴリーの魅惑の車が、珍しく東京都心の台場でレースを行う格別な地域と期間が抽出されており、撮影スポットと言える地域が良く抽出できている点で、提案法の有効性を実証している。

また、各地域の時間軸に沿った詳細を調べるため、2010 年 365 日の日ごとに、どの程度の写真が撮影されているかを色で示したのが図 3 である。横軸は日、縦軸は Rank 1 ~ Rank 10 までを黒の直線で区切り、縦に並べている。ただし、日ごとの写真数は、0 枚は白、1 ~ 50 枚を灰色、50 枚以上を黒で表現した。図 3 から分かるように、提案法で抽出されたホット撮影期間を含む主要撮影地域は、年間の写真総数が少ないことは一目瞭然である。また、表 2 の提案法が抽出した撮影期間を図 3 の各 Rank ごとの下部に横軸の日数に沿って赤線で示した。図 3 から、提案法が抽出した地域には、普段とは逸脱して顕著に撮影数が増えた、時期と相関した旬の期間 (黒が含まれる期間) が存在することがわかる。また、赤線で示した提案法で抽出されたホット撮影期間がその旬の期間と良く対応し、抽出性能の良さを実証している。また、その旬の期間以外、撮影がほとんど行なわれていないため、その旬の期間は、その地域にとって格別な期間である可能性が示唆される。そこで、抽出されたホット撮影スポット候補を抜粋し、詳細を考察する。

#### 4.3 考 察

提案法のランキングトップとなった醍醐寺は、世界遺産であり、春の桜や秋の紅葉を含め、四季を通じて見どころがあり、多くの人々が訪れる観光スポットであると考えられる。しかし、図 3 を見る限り、1 年のうち写真が撮影されている時期は 3 月の終わりから 4 月中旬に

かけた数週間に集中している。この時期、桜は日本全国各地で開花するため、桜を觀賞する地域としては特別ではない。しかし、醍醐寺は、豊臣秀吉が贅をつくし、各地の有名な桜を醍醐寺に集め、かわづ桜、しだれ桜、ソメイヨシノ、山桜、八重ザクラ、大紅しだれ、大山桜などを一箇所で見る事ができる桜の名所としては格別の場所であると言える。図4は、醍醐寺の地域に含まれるホットスポット写真の例であるが、多種の桜が含まれる。醍醐寺では、この時期以外に写真の撮影がほとんどないことから、この時期にこそ、撮影者の心をつかみ、シャッターを押したくなる格別の場所であることを教えてくれていると見なせる。

また、ランキング6位の愛知県豊田市小原は、都市から離れた奥地でありながら、図3のように、ほぼ11月中旬にのみ撮影が行なわれている。小原には、四季桜という4月と10月末の年に2回咲くめずらしい桜が植えられている。四季桜は4月にも咲くが、図3から、4月には撮影が行なわれていない。つまり、単に桜が見れるだけでは、この奥地で桜を撮影しに訪れる人はいないことを示唆しており、10月末から11月中旬に紅葉と桜を同時に見るのできる格別のこの時期にこそ、撮影者をこの地域に呼び寄せると推測される。これらの例は、提案手法が撮影者の意見をくみ取り、特別なホット撮影スポットを発見できた好例であると言え、この点においても提案法の有効性が実証された。

## 5. ま と め

本研究では、数多くの人々が撮影した、撮影場所を示す地理情報や撮影時間の情報が付与された大量の写真データを用いて、実世界の空間に局在する主要撮影地域と、時間軸上に局在し、他の地域と比較して顕著にその地域で写真が撮影されているホット撮影期間とのペアをホット撮影スポットと呼んで、その局在する実時空間の抽出問題に取り組んだ。そして、ホット撮影スポットの数理モデルと、その効率的な抽出法を提案し、実データを用いた実験により、提案法の有効性を実証した。今後は、応用を検討し、さらなる効率性や拡張性を求め、ホット撮影スポットの抽出問題の優れた解法を探索する。そのさらなる発展に向け、ホット撮影スポットの自動抽出という新たな問題を考え、有効かつ効率的な手法を提案した本研究は、その重要な第一歩であると言える。

謝辞 本研究の一部はJSPS 科研費 23650113 の助成を受けたものである。

## 参 考 文 献

- 1) 垂水浩幸：ソーシャルメディアと実世界，情報処理，Vol.51, No.7, pp.782–788 (2010).
- 2) Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event

- Detection by Social Sensors, *Proceedings of the 18th International Conference on World Wide Web*, pp.851–860 (2010).
- 3) 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).
  - 4) 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).
  - 5) 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).
  - 6) Crandall, D.J., Backstrom, L., Huttenlocher, D. and Kleinberg, J.: Mapping the world's photos, *Proceedings of the 18th International Conference on World Wide Web*, pp.761–770 (2009).
  - 7) Yin, Z., Cao, L., Han, J., Zhai, C. and Huang, T.: Geographical Topic Discovery and Comparison, *Proceedings of the 20th International Conference on World Wide Web*, pp.247–256 (2011).
  - 8) 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).
  - 9) Arase, Y., Xie, X., Hara, T. and Nishio, S.: Mining People's Trips from Large Scale Geo-tagged Photos, *Proceedings of the 18th International Conference on Multimedia*, pp.133–142 (2010).
  - 10) Yin, H., Lu, X., Wang, C., Yu, N. and Zhang, L.: Photo2Trip: an interactive trip planning system based on geo-tagged photos, *Proceedings of the 18th International Conference on Multimedia*, pp.1579–1582 (2010).
  - 11) Lu, X., Wang, C., Yang, J.-M., Pang, Y. and Zhang, L.: Photo2Trip: generating travel routes from geo-tagged photos for trip planning, *Proceedings of the 18th International Conference on Multimedia*, pp.143–152 (2010).
  - 12) Swan, R. and Allan, J.: Automatic generation of overview timelines, *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.49–56 (2000).
  - 13) 福井健一，斉藤和巳，木村昌弘，沼尾正行：自己組織化ネットワークによる動的クラスタの可視化編纂，人工知能学会論文誌，Vol.23, No.5, pp.319–329 (2008).