

## 集合知に基づいた ポピュラー撮影スポットに関する旬シーズンの可視化

熊野雅仁<sup>1)</sup>(正会員) 岩渕聡<sup>2)</sup>(非会員) 小関基徳<sup>2)</sup>(非会員)  
小野景子<sup>1)</sup>(非会員) 木村昌弘<sup>1)</sup>(非会員)

1) 龍谷大学 理工学部 2) 龍谷大学大学院 理工学研究科

## Visualization of Attractive Periods for Popular Photo Spots Based on Collective Wisdom

Masahito Kumano<sup>1)</sup> Satoshi Iwabuchi<sup>2)</sup> Motonori Koseki<sup>2)</sup>  
Keiko Ono<sup>1)</sup> Masahiro Kimura<sup>1)</sup>

1) Faculty of Science and Technology, Ryukoku University

2) Graduate School of Science and Technology, Ryukoku University

kumano @ rins.ryukoku.ac.jp

### 概要

本研究は、ソーシャルメディアデータを用いて、人々の集団的行動を分析するための可視化問題に取り組む。近年、Flickrなどの写真共有サイトやFacebookなどのソーシャルメディアにおいて大量に蓄積されている、GPSに基づく撮影位置や、撮影時間、撮影者などの、デジタル写真に付随したメタ情報を利用し、ユーザ群の実際の行動をうまく集約することで、観光支援システムの構築に応用する研究が注目されており、多くの人々が集まる空間上の人気スポットを地図上に可視化した研究がある。しかし、日本有数の観光名所である清水寺では、春の桜、秋の紅葉の時期に、より多くの人々が集まるように、各撮影スポットには独自の魅力的な旬シーズンが存在し得る。また、旬シーズンを持つ撮影スポットは、各地に多数存在し、人々が集まりやすい時期が異なり得ると考えられる。我々は、観光支援システムをより洗練させるため、旬シーズンの魅力度を定量化し、魅力度を視覚化した複数の撮影スポット群を一度に鳥瞰し、パースト性の観点から視覚的に時空間的特徴を分析、理解させ得る、効果的な可視化法を提案する。日本で撮影されたFlickrの写真データを用いることで、提案する可視化分析法の効果を実証する。

キーワード ビッグデータ、集合知、可視化、視覚的データマイニング、時空間データマイニング

### Abstract

We address the visualization problem for analyzing collective human behavior using Social Media data. Recently, attention has been devoted to constructing sightseeing guide systems that exploit the information revealed by the collective behavior of users in photo-sharing websites such as Flickr and Facebook, where photos are annotated with GPS locations, time-stamps, photographers, etc. Previous work discovered popular photo spots from a large number of geo-tagged photos, and visualized them on maps. For each popular photo spot, we focus on its burst season as a candidate for its attractive period. For instance, Kiyomizu which is a main sightseeing spot of Kyoto in Japan has a number of the burst seasons that attract many visitors in the season of cherry blossoms or autumn leaves. By effectively visualizing burst seasons, we aim to increase the sophistication of a sightseeing guide system based on collective wisdom. We quantitate degree of the attractiveness and propose a visualization method that visualize the collective behavior and spatio-temporal information all at once, and can effectively analyze and compare the popular photo spots in terms of burstiness. Using Japanese Flickr data, we demonstrate that the proposed visual analysis method is effective.

**Keyword** Big data, Collective wisdom, Visualization, Visual data mining, Spatio-temporal data mining

## 1 はじめに

GPS 受信機能付カメラの普及と、Flickr<sup>1</sup>など、写真共有サイトの繁栄により、GPS に基づく撮影位置 (Geo-tag) や、撮影時間、撮影者などのメタ情報が付随した膨大な写真が Web 空間に蓄積され続けており、どこで、いつ、誰が写真を撮影したかを把握することのできる大量の写真データがビッグデータとして利用可能となっている。人々は、外出するとき、視覚的に興味を抱いた対象に出くわすと写真を撮影する傾向があり、撮影者が、多種・多様な嗜好に基づいて、実世界の魅力や事象・事物を厳選して捉え、社会に情報提供する有能なフィルタ付ソーシャルセンサ [1] と見なせる可能性があるため、メタ情報付き写真群から撮影者の行動をうまく集約し、それらに基づいて新たな価値を創造する研究が注目されている。新たな価値が創造された例として、Crandall ら [2] は、大量の Geo-tag 付き写真を用いて人々の実際の撮影行動を集約し、代表写真を地図上にマッピングし、ランドマークのある人気撮影スポットの所在を地図上で視認可能にした。この研究は、魅力的な場所の探索 [3, 4] や、主要な撮影スポット (地域・場所) を抽出し得ることから、観光に応用する研究 [5, 6, 7, 8] に波及し、この他にも、地域ごとの地理的トピックを抽出し、地域間を比較して文化的特徴の違いを分析する研究 [9] にも波及している。ただし、これらは、基本的に空間に主眼がある。一方、特に日本の場合、人気撮影スポットには、空間的に限られた人気のある地域という一面以外に、例えば、春の桜、秋の紅葉など、四季に応じた人気のある時期、つまり旬シーズンが存在し得る。ただし、旬シーズンを持つ人気撮影スポットには、フェスティバルやイベントなど、必ずしも四季とは関係しない事例も存在し得る。また、それらの事例は、撮影スポットごとに、時期や期間が異なり得るだけでなく、同じスポットでも、異なる時期に異なる事例で人々が集まる場合も考えられ得る。つまり、各人気撮影スポットには、時期も違えば、期間、数も異なる独自の旬シーズンが存在し得ると考えられる。旬シーズンが明らかになれば、各撮影スポット独自の魅力を、より明瞭化し得るため、観光への応用が期待される。したがって、各撮影スポット独自の旬シーズンを抽出する有効な手法が必要となる。

さらに、観光への応用に向けては、一つの撮影スポットに関する旬シーズンを捉えるだけでなく、異なる撮影スポット間の旬シーズンを比較分析できることが望ましい。例えば、京都には、桜や紅葉など、複数の時期が楽しめる撮影スポットが近隣にいくつも存在する。さらに、人々は、より見応えのある対象を求める傾向があるので、近隣スポットと比較することで、集合知の観点から、高い旬の割合を持つ名所を見い出せる可能性がある。また、複数のスポットを比較する上で、旬の割合だけでなく、各シーズンの時期、期間

など多種の特徴を同時に比較できれば、各撮影スポットの特色がより明瞭になる可能性がある。近年、気象データを対象として、関連する多種情報を同時に可視化する研究 [10] が行われているが、この研究においては、情報を読み取り得る視認性が重要となる。近年、情報を読み考えるための可視化 [11] や視覚的にデータマイニングを行う Visual Data Mining (VDM) [12] などに始まる、新たな発見や知見の獲得を促すための効果的な可視化法が注目されつつある。本研究においては、数多く存在する、異なる撮影スポット間の同時比較を可能とするため、撮影スポット群を一望視しながら多種情報を比較し得る、効果的な可視化法に着目する。また、視認性を高めるためには、一望視するだけでなく、注目する空間スケールに応じて、対話的に見方を変更できる環境を構築することが望ましい。本研究では、3DVDM [13] のように、3D 情報空間へ没入して対話的に VDM の支援を行う環境に注目しており、特に、地図空間上に撮影スポットを配置することで、地理的關係も視認できる環境構築法に着目する。

そこで、本研究では、観光支援システムをより洗練させるため、撮影スポット独自のシーズンを抽出する手法を与え、旬の割合を定量化し、空間と時間および撮影者行動に関連する以下の五つの分析観点

- 1) 各スポットにいくつのシーズンが含まれるか、
- 2) 各シーズンはどの時期か、
- 3) 各シーズンはどの程度のバースト性を示すか、
- 4) 各スポットで、年間、どの程度人が訪れているか、
- 5) スポット同士はどのような空間的關係にあるか、

に関する多種情報を 3D 地図空間において同時に視認可能とする、効果的な可視化法を提案する。また、Flickr から収集した実データを用いて、集合知的集約に基づいたシーズン抽出の質の良さや可視化結果を評価し、提案法の有効性を示す。そこで、まず、2章で入力データからポピュラー (人気) 撮影スポットを抽出する手法を述べ、3章でシーズンの抽出法と旬の割合の定量化法を述べ、提案可視化法について述べる。次に、4章でシーズン抽出の品質と可視化実験結果の評価を行い、議論を行って、5章でまとめる。

## 2 ポピュラー撮影スポット抽出

大量のメタ情報 (Geo-tag および撮影者 ID) 付き写真群を用いて実際の人々の写真撮影行動を集約し、集合知的観点からポピュラー撮影スポットを抽出するため、与えられた写真データを撮影者数の観点から捉えなおし、個数が未知のポピュラー撮影スポット群を自動抽出する手法を示す。

### 2.1 入力データ

正の整数  $\mathcal{J}$  に対して、1年の期間  $\mathcal{J}$  内で撮影された写真データ全体の集合を  $\mathcal{D}_0$  とする。写真は、撮影者により、同じ地点で1枚撮影する者もいれば、多数の撮影を行う者など、個人差があると考えられる。

<sup>1</sup><http://www.flickr.com/>

本研究では、集合知的観点から撮影者の人数を重視するため、個人差を吸収し、緯度と経度の2次元平面を離散化し、最小矩形領域ごとに、1人の撮影者による写真を1枚抽出する正規化を行う。このとき、最小矩形領域内の写真数は、その領域内で撮影を行った人数に相当することを注意しておく。この離散正規化を  $\mathcal{D}_0$  に適用して得られる写真集合を、

$$\mathcal{D}_1 = \{d_n; n = 1, \dots, N\}$$

とする。ここに、各写真データ  $d_n$  には、位置情報  $x_n$  と、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 だけでなく、Panoramio、Facebook、Instagram など、多くのソーシャルメディアから、このような写真データが大量に収集可能であることを注意しておく。

## 2.2 抽出法

大量の Geo-tag 付き写真データ  $\mathcal{D}_1$  が与えられたとき、撮影位置の密集する数が未知の地域をポピュラー撮影スポット群として自動抽出する手法を示す。ここで、 $d$  次元 Euclid 空間  $\mathbf{R}^d$  上の点群を、ある確率分布に従う標本集合  $S = \{s_n\}_{n=1, \dots, N}$  と考えたとき、任意の点  $s \in \mathbf{R}^d$  における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

$$\hat{p}(s) = \frac{1}{N} \sum_{n=1}^N \frac{1}{h_n^d} G(\| (s - s_n) / h_n \|^2), \quad (s \in \mathbf{R}^d) \quad (1)$$

を考える。ここに、 $\|\cdot\|$  は  $\mathbf{R}^d$  の Euclid ノルム、 $G(z)$  はカーネル関数である。ここで、 $G(z)$  は、Uniform、Epanechnikov や Gaussian カーネルなどを利用する。また、 $h_n (> 0)$  は、データの各位置  $s_n$  ごとに存在する確率密度関数のバンド幅を規定するパラメータである。この確率密度関数の極大点を発見する効率的な手法として、Mean Shift 法 [14] がある。また、バンド幅を自動的に決定する手法に Variable Bandwidth Mean Shift (VBMS) [15] があり、バンド幅  $h_n$  を  $h$  に固定化する Fixed Bandwidth Mean Shift (FBMS) も用いられる。

Crandall ら [2] は、写真データ集合  $\mathcal{D}_1$  から主要な撮影地域  $\mathcal{R}_k$ , ( $k = 1, \dots, K$ ) を抽出する手法として、緯度と経度に基づく2次元の点群  $\{x_n\}$  を対象として Mean Shift クラスタリングを適用している。ただし、主要な撮影地域は、例えば、街角レベルから都道府県、州、国規模など、様々なスケールが考えられるため、最適なサイズを容易に決定できない問題がある。Crandall らは、カーネル幅  $h_n$  を  $h$  と固定し、対象とする問題のスケールに応じて Metropolitan-scale( $h=100\text{km}$ ) と Landmark-scale( $h=100\text{m}$ ) として実験を行っている。

本研究では、観光への応用に向け、徒歩圏内の多くの撮影者が集まる地域をポピュラー撮影スポットと想定し、抽出する上で、Crandall らの Landmark-scale を参考として  $h=100\text{m}$  に固定し、Crandall らと同様に、FBMS 法を用いる。ただし、極大値が比較的小さいものについては、主要な撮影地域とは考えないことにする。 $G(z)$  は Epanechnikov カーネル

$$G(z) = \begin{cases} 1 - z & (|z| \leq 1) \\ 0 & \text{otherwise} \end{cases}$$

を用いる。

## 3 提案法

### 3.1 旬シーズンの抽出法

#### 3.1.1 シーズン抽出

FBMS 法に基づいて抽出された地域  $\mathcal{R}_k$ , ( $k = 1, \dots, K$ ) において、各  $\mathcal{R}_k$  は、独自のシーズンを持つと考える。ただし、存在するシーズンの数や時期が未知であるだけでなく、期間(日数)も未知であり、複数のシーズンが存在し得る場合、それぞれの期間は短い場合や長い場合があることも予想される。

ここで、シーズンを自動抽出するため、撮影日の集合  $\mathcal{T}_k = \{t_n; x_n \in \mathcal{R}_k\}$  を対象とし、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を考える。各写真  $d_n$  は、独自のカーネル幅  $h_n$  を持つと想定し、 $h_n (> 0)$  を適応的に決定できる VBMS 法 [15] を用いることで、各  $\mathcal{T}_k$  に対し、クラスタリングを行って、 $\mathcal{R}_k$  ごとに、 $M_k$  個のシーズンの集合  $\{\mathcal{J}_{k,1}, \dots, \mathcal{J}_{k,M_k}\}$  を出力する。ここに、各  $\mathcal{J}_{k,m}$  は、閉区間  $\mathcal{J}_{k,m} = [t_{k,m}^{min}, t_{k,m}^{max}]$  である。

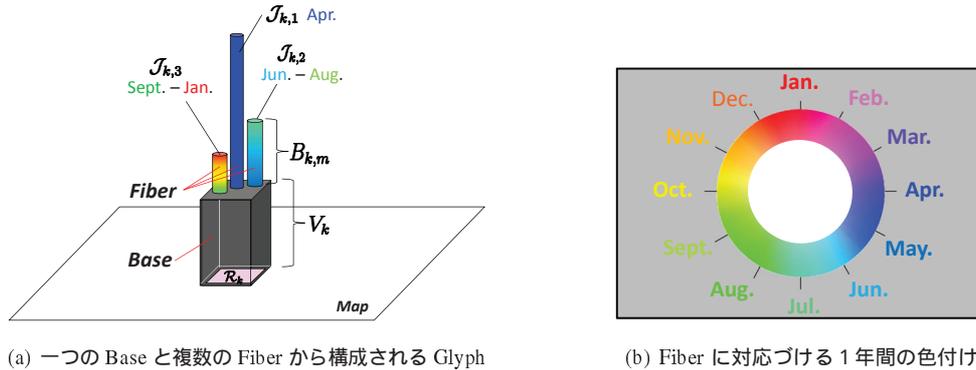
#### 3.1.2 パースト度

本研究では、より短い期間に撮影行動が頻発するほど、集合知的観点において旬の度合いが高いシーズンであると考え、旬の度合いを定量化する観点として、パースト度に着目した指標

$$B_{k,m} = \frac{V_{k,m}/V_k}{|\mathcal{J}_{k,m}|/|\mathcal{J}|} \quad (2)$$

を提案する。ここで、 $V_{k,m}$  はシーズン  $\mathcal{J}_{k,m}$  内で  $\mathcal{R}_k$  に訪問した撮影者数を表わし、 $V_k$  は、期間  $\mathcal{J}$  における  $\mathcal{R}_k$  を訪れた撮影者数を表わしている。また、 $|\mathcal{J}_{k,m}|$  は  $\mathcal{R}_k$  における  $m$  番目のシーズンの期間日数、 $|\mathcal{J}|$  は全期間  $\mathcal{J}$  の日数を表わしている。

$B_{k,m}$  は、期間全体  $\mathcal{J}$  に対し、シーズン  $|\mathcal{J}_{k,m}|$  が短いほど大きく、 $\mathcal{R}_k$  の撮影者数  $V_k$  に対し、注目するシーズン  $\mathcal{J}_{k,m}$  の撮影者数  $V_{k,m}$  が多いほど大きくなるため、短い期間であるほど、また、多くの撮影者が含まれるほど大きい値を示す。本研究では、 $\mathcal{J}_{k,m}$  の  $B_{k,m}$  が大きいほど、旬の度合いが高いシーズンと見なす。



(a) 一つの Base と複数の Fiber から構成される Glyph

(b) Fiber に対応づける 1 年間の色付け

図 1: 一つのポピュラー撮影スポットに対応する Glyph

### 3.2 可視化法

観光への応用や地域分析に向けて、五つの観点

- 1) 各スポットにいくつのシーズンが含まれるか、
- 2) 各シーズンはどの時期か、
- 3) 各シーズンはどの程度のペース性を示すか、
- 4) 各スポットで、年間、どの程度人が訪れているか、
- 5) スポット同士はどのような空間的關係にあるか、

から、ソーシャルセンサの集団的行動に関する視覚的な分析を可能にするため、抽出された数多くのポピュラー撮影スポット  $\mathcal{R}_k$  と、 $\mathcal{R}_k$  が占める領域の範囲、また、各  $\mathcal{R}_k$  へ年間に訪れる撮影者数  $V_k$ 、さらに、各  $\mathcal{R}_k$  独自のシーズン  $\mathcal{J}_{k,m}$  の数と、各  $\mathcal{J}_{k,m}$  の時期や期間、各シーズンの旬の割合  $B_{k,m}$  など、多種ある時空間情報を同時に視認可能とする効果的な可視化法を与える。

図 1(a) に、一つのポピュラー撮影スポット  $\mathcal{R}_k$  を可視化したグリフ (Glyph) の一例を示す。各グリフは、まず、各  $\mathcal{R}_k$  に含まれる写真群の撮影位置を包含する最小の矩形領域を底辺とした四角柱が地図上に配置される。この四角柱をベース (Base) と呼ぶ。このベースの底辺の大きさから、 $\mathcal{R}_k$  が占める領域の大きさが視認可能となる。また、四角柱の高さを  $V_k$  とすることで、 $\mathcal{R}_k$  へ訪れる年間の撮影者数が視認可能となる。

一方、Base の上部には、各  $\mathcal{R}_k$  で抽出されたシーズンに関する情報を円柱として可視化する。この一つの円柱は、抽出された一つのシーズン  $\mathcal{J}_{k,m}$  に対応し、ファイバー (Fiber) と呼ぶ。ファイバーの数によって、各  $\mathcal{R}_k$  から抽出されたシーズン数が視認可能となる。また、ファイバーの高さに  $B_{k,m}$  を割り当てることによって、高いファイバーほど旬の割合が高いシーズンであることが視認可能となる。さらに、各ファイバーには、図 1(b) に示す色を割り当てる。この色は、1 年間の各日に HSV 色相環に基づく色を恣意的に割り当てたものであり、ファイバーの下端を  $\mathcal{J}_{k,m}$  の  $t_{k,m}^{min}$ 、上端を  $t_{k,m}^{max}$  とすることで、ファイバーが単色であるほど  $\mathcal{J}_{k,m}$  の期間が短く、多色であるほど長い期間であることや、 $\mathcal{J}_{k,m}$  のおおよその時期と期間を同時に視認可能とすることができる。この可視化法により、図 1(a)

の  $\mathcal{R}_k$  は、三つのシーズンを持ち、図 1(b) を参照することにより、 $\mathcal{J}_{k,1}$  は 4 月頃の短いシーズン、 $\mathcal{J}_{k,2}$  は 6 月頃から 8 月頃のシーズン、 $\mathcal{J}_{k,3}$  は 9 月から 1 月頃と比較的長いシーズンであることや、特に  $\mathcal{J}_{k,1}$  シーズンの旬の割合が他のシーズンと比較して相対的に高いことが容易に視認可能となり得ることもわかる。

また、抽出された数多くの  $\mathcal{R}_k$  を一度に可視化し、全体を一望可能とすることで、複数の  $\mathcal{R}_k$  が密集する地域や、疎な地域がわかるだけでなく、 $\mathcal{R}_k$  が密集する地域において、 $\mathcal{R}_k$  同士の空間的な位置関係や、年間に訪れる撮影者数の相対的な差がわかるため、ベースとファイバーを総合的に視認することで、五つの観点を満たした視覚的分析を可能とすることがわかる。さらに  $\mathcal{R}_k$  同士の旬シーズンが同じもしくは違う傾向を示すか、どの旬シーズンが多いか、旬シーズンの旬の割合の比較など、地域のシーズンに関する特色や傾向を多面的に視認しながら分析し得ることがわかる。

尚、Glyph の可視化は、OpenGL を用いて実装し、キーボードを用いて、3D 地図空間内を自由に移動できるように、システムを構築した。

## 4 実験

### 4.1 実験設定

本研究は、特に旬シーズンの可視化に関する効果を検証するため、日本国内で撮影された Geo-tag 付き写真データを対象として提案法の有効性を示す。このため、写真共有サイト Flickr から、日本で撮影されたことを示す WOEID (Where On Earth ID): 23424856 を持つ 2010 年の 1 月 1 日から 12 月 31 日までの位置・時間・ユーザ ID 情報付き写真データを収集したところ、 $|\mathcal{D}_0| = 548,922$  枚の写真データ集合  $\mathcal{D}_0$  が得られた。次に、2.1 節で述べたように、 $\mathcal{D}_0$  に対し、集合知的観点から個人差を吸収するため、最小矩形領域ごとに、一人の撮影者が何枚撮影していても 1 枚と数える正規化を行った。ただし、最小矩形領域のサイズについては、例えば桜や紅葉を撮影する際、同じ木々を対象とする場合でも、撮影者の嗜好により、別の地点 (視点) から

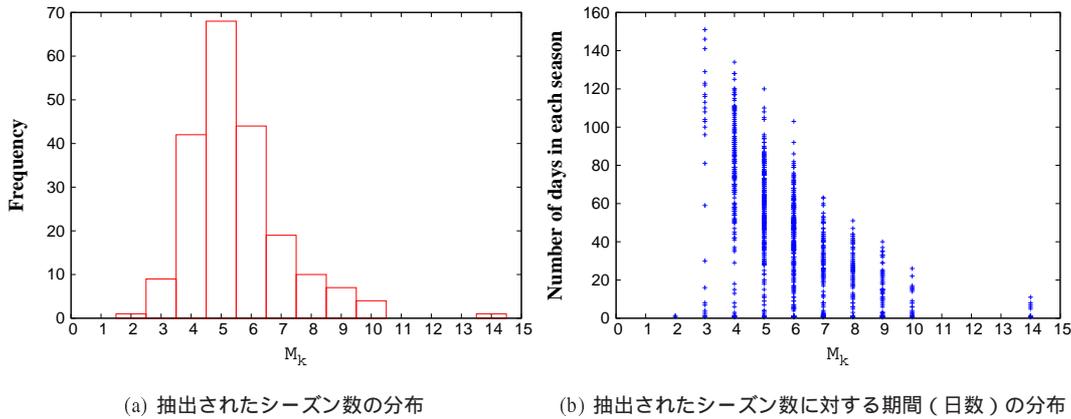


図 3: 提案法による抽出されたシーズン数に関する分布

撮影する場合が考えられる．この地点の違いを嗜好の違いとして区別するため，異なる最小領域は，異なる撮影地点であると見なし，最小矩形領域の1辺を10mとして実験を行った．正規化の結果， $|\mathcal{D}_1|=162,933$ 枚の写真データ集合  $\mathcal{D}_1$  が得られた．本研究では， $\mathcal{D}_1$  を Flickr dataset と呼ぶ．

Flickr dataset に対し， $h=100\text{m}$  として FBMS 法を適用したところ， $K'=24,954$  個の地域が抽出されたが，撮影者数の多い地域に焦点を当てるため， $V_k > 100$  を満たす地域に着目したところ，可視化の対象とするポピュラー撮影スポット数は， $K=205$  となった．

## 4.2 シーズン抽出結果の評価

日本においてシーズンを考える場合，最も単純には四季に分割する方法が考えられる．提案法により自動抽出されたシーズンは，四季とは必ずしも一致せず，より短いもしくは長い期間となり得る．四季に分割するよりも，提案法により抽出されるシーズンの有効性を検証するため，まず，抽出されたシーズン数に関する傾向を示した後，シーズン抽出に関する品質の観点から評価を行う．さらに，抽出された各シーズンのバースト度の傾向についても調べる．

### 4.2.1 抽出されたシーズン数の結果

京都は，四季折々の観光名所が多数含まれる地域として有名であるが，京都で最も観光客が多く，ほぼ年中人気のある撮影スポットとして，清水寺が知られている．図 2 に，清水寺に着目したシーズン抽出例を示す．図 2 の横軸は，1 年の各日を表わしており，縦軸のうち，黒の棒グラフで表されたヒストグラムは撮影者数（訪問者数），赤で表された曲線は，撮影者数の分布から推定された確率密度関数 (pdf: probability density function) を表わしている．また，清水寺では， $\mathcal{J}_{k,1}$  から  $\mathcal{J}_{k,6}$  までの  $M_k=6$  個のシーズンが抽出され

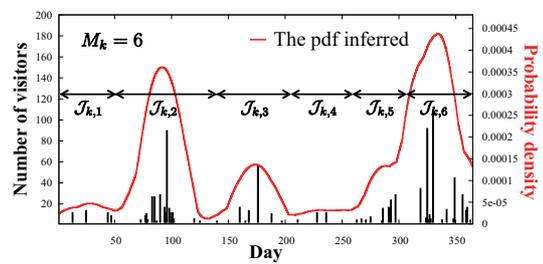


図 2: シーズン抽出の例”清水寺”(2010)

ており，図 2 中，両矢印で示された区間は，推定された各  $\mathcal{J}_{k,m}$  の期間を表わしている．

ここで， $K=205$  個の主要撮影スポット  $\mathcal{R}_k$  について，抽出されたシーズン数  $M_k$  の傾向を調べるため，横軸に  $M_k$ ，縦軸に  $M_k$  個のシーズンが抽出された主要撮影スポット数をとった棒グラフを図 3 (a) に示す．図 3 (a) より，シーズン抽出数は，4 つ（四季）ではなく，5 つが最も多く，最も少ない 2 から，年間月数 12 を上回る 14 まで分布していることから，それぞれの  $\mathcal{R}_k$  が独自のシーズン数を持ち得る結果となった．

### 4.2.2 抽出されたシーズン期間の結果

次に，図 2 より，清水寺は，ほぼ一年に渡って撮影が行われているが，撮影者の訪問数の違いに基づいて抽出された各シーズン  $\mathcal{J}_{k,m}$  の期間長が，それぞれ異なる様子が伺える．ここで，抽出されたシーズンの期間長に関する傾向を調べるため， $M_k$  に対する，抽出されたシーズンの期間（日数）の分布を図 3 (b) に示す．四季に区分する場合，単純には，一つの季節が約 3 か月となり，日数としては 90 日程度になるが，図 3 (b) より， $M_k$  が 3 から 6 となる  $\mathcal{J}_{k,m}$  では，最大で 150 日程度まで，90 日より長い期間が多数存在することや，全体的に，90 日より短い期間が多数存在することもわかる．つまり，提案法で抽出された

$\mathcal{J}_{k,m}$  には、異なる多様な期間が存在し得るという結果となった。

#### 4.2.3 抽出されたシーズンの品質評価

各  $\mathcal{R}_k$  は、独自のシーズン数や期間長を持ち得るとしても、提案法で抽出されたシーズンが、人々の撮影行動をうまく捉えているとはかぎらない。そこで、抽出されたシーズンの有効性を検証する評価法を考える。

もしシーズンがうまく抽出されていれば、シーズン間に比べ、シーズン内で撮影された写真同士が似たような対象を捉えている傾向が高いと考えることができる。ここで、各シーズンを写真群からなるクラスタと捉えたとき、クラスタ内の写真間距離は近く、クラスタ間の写真間距離は遠くなるのがより望ましいと考える。つまり、クラスタ間距離に対するクラスタ内距離が小さいほど、抽出されたシーズンの品質が良いと見なす。そこで、本研究では、 $\mathcal{R}_k$  ごとに、Davies-Bouldin 基準 [16]

$$DB_k = \frac{1}{M_k} \sum_{i=1}^{M_k} \max_{i \neq j} \left( \frac{\sigma_{k,i} + \sigma_{k,j}}{D(\mathcal{I}_{k,i}, \mathcal{I}_{k,j})} \right)$$

に基づいて、抽出されたシーズンの品質に関する評価を行う。ここで、 $\mathcal{I}_{k,m}$  は、 $\mathcal{R}_k$  ごとに、 $M_k$  個抽出された各シーズン  $\mathcal{J}_{k,m}$  に撮影された写真集合を表わす。また、 $\sigma_{k,m}$  は、 $\mathcal{I}_{k,m}$  内の任意の写真画像ペア  $v, w \in \mathcal{I}_{k,m}$  に関する、総当たりによる写真間距離の平均

$$\sigma_{k,m} = \sum_{v \neq w} \text{dist}(v, w) / \binom{|\mathcal{I}_{k,m}|}{2}$$

を表わし、 $\text{dist}(v, w)$  は、写真間距離、 $D(\mathcal{I}_{k,i}, \mathcal{I}_{k,j})$  は、クラスタ  $\mathcal{I}_{k,i}$  と  $\mathcal{I}_{k,j}$  間の写真間距離の平均

$$D(\mathcal{I}_{k,i}, \mathcal{I}_{k,j}) = \frac{1}{|\mathcal{I}_{k,i}| |\mathcal{I}_{k,j}|} \sum_{i \neq j} \sum_{v \neq q} \text{dist}(v, q)$$

を表わしている。ただし、 $v \in \mathcal{I}_{k,i}, q \in \mathcal{I}_{k,j}$  である。写真間距離  $\text{dist}(v, q)$  を計算するための画像特徴としては、基本的な SURF [17] を用いて BoVW (Bag of Visual Words) で表現する<sup>2</sup>。尚、BoVW 表現は、 $\mathcal{R}_k$  ごとに、含まれる写真すべてを用い、SURF 特徴を一般的に用いられる 1000 次元でベクトル量子化し、量子化された各特徴を visual word として、各写真内で visual word が出現する頻度を求め、各 visual word の頻度を

<sup>2</sup>テキスト解析においては、文書を単語の頻度情報で表現する BoW (Bag-of-Words) 表現に基づいて、文書コーパスを統計解析することが盛んに行われている。画像解析においても、近年、スケール、回転および照明変化に不変な局所特徴量である SIFT (Scale Invariant Feature Transform) や SURF (Speeded-up robust features) が発見され、それらを用いて、画像を visual word の頻度情報として表現 (BoVW 表現) し、テキスト解析と同様な手法を用いて画像コーパスを統計解析することが盛んになりつつある [18]。本論文では、SURF 特徴量による BoVW 表現に基づいて分析する。

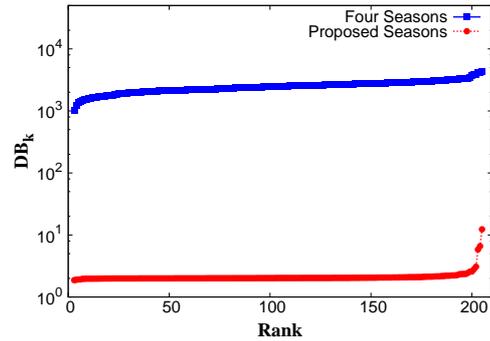


図 4: シーズン抽出の品質比較

ピンとするヒストグラムで表現した。また、シーズン内およびシーズン間における写真画像の類似・非類似傾向をより明瞭化するため、ヒストグラムが異なるほど距離が顕著に大きくなり、ヒストグラム距離として正確さで定評のある Chi-square 距離を用いた [19]。

本研究では、提案法で抽出されたシーズンと、最も単純に 1 年を四季に分けた場合とで、Davies-Bouldin 基準によるクラスタの品質を比較することにより、提案法の有効性を検証する。比較法として、1 年を四季に分ける方法としては、 $\mathcal{R}_k$  ごとに、(春:3 月～5 月、夏:6 月～8 月、秋:9 月～11 月、冬:12 月～2 月) に分割した<sup>3</sup>。図 4 は、 $K=205$  個の  $\mathcal{R}_k$  ごとに  $DB_k$  を計算し、 $DB_k$  を昇順にランキングしたとき、横軸をランク  $r$  とし、縦軸を  $r$  位の  $DB_k$  として、四季に分けた場合と、提案法を比較した結果である。 $DB_k$  は、値が小さいほど抽出されたシーズンの品質が良いと評価されるため、提案法は、シーズンを四季で抽出した場合と比べ、格段に品質が良い結果を示していることから、提案法のシーズン抽出に関する有効性が示唆された<sup>4</sup>。

#### 4.2.4 パースト度の傾向

抽出されたシーズン  $\mathcal{J}_{k,m}$  ごとに、提案法によって句の割合を示すパースト度  $B_{k,m}$  が算出される。算出された  $B_{k,m}$  の傾向を調べるため、横軸を  $B_{k,m}$  としたとき、 $B_{k,m}$  以上の値を持つ  $\mathcal{J}_{k,m}$  の存在確率を縦軸とした結果を図 5 に示す。図 5 より、 $B_{k,m}$  は、変曲点が 1.5 程度であり、0.1 から 100 を超えるものまで、多様なパースト度を示すシーズンの存在が示唆された。

<sup>3</sup>本研究では、同じデータセットで品質を検証するため、冬の 1 月と 2 月の写真データは翌年ではなく、同じ年の 1 月と 2 月の写真データを用いた。

<sup>4</sup>提案法は、写真の撮影位置の情報から抽出された主要撮影スポットに対し、そのシーズンを撮影日の情報のみから検出しているため、同じシーズンに属する写真の方が異なるシーズンに属する写真よりも、画像情報において類似しているとは限らないことに注意しておく。ところで、画像情報の類似性の自動判別は一般に限界があり得るので、被験者実験によるシーズン検出の詳細評価は今後の重要な課題である。

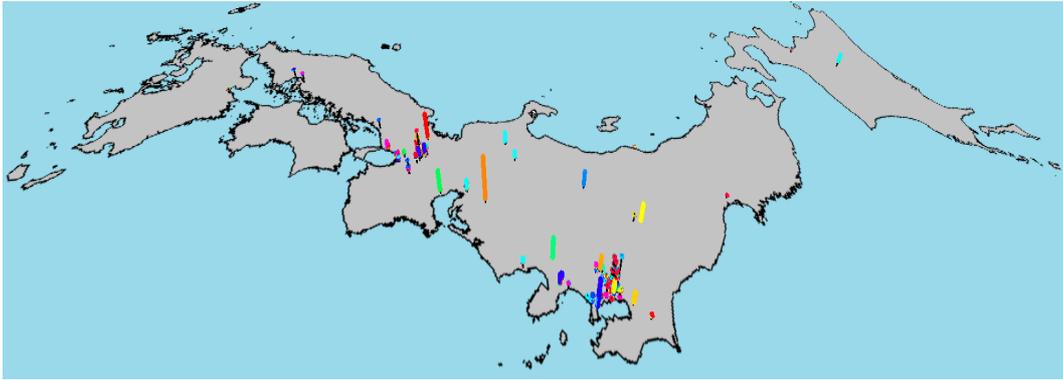
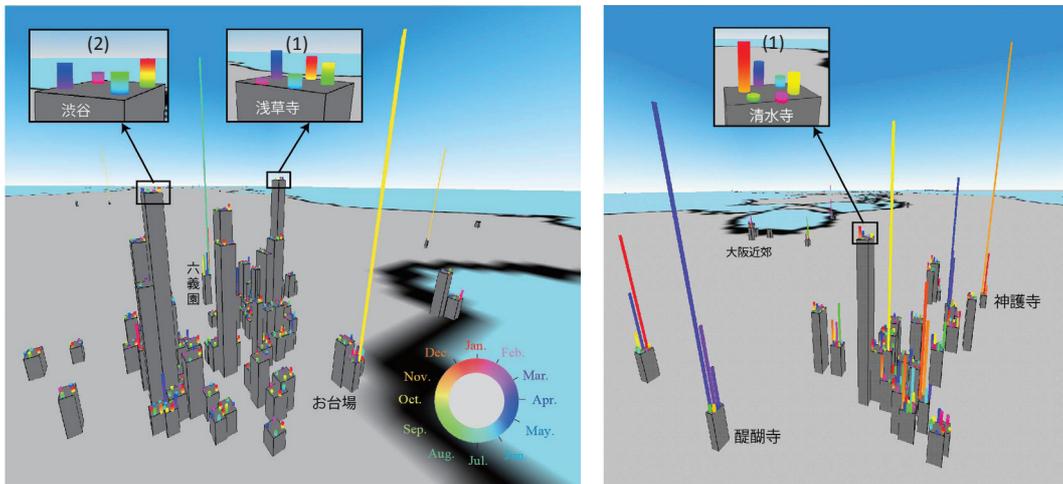


図 6: 提案法に基づいてポピュラー撮影スポットを日本地図上に可視化し，鳥瞰視した結果



(a) Tokyo region

(b) Kyoto region

図 7: 東京近郊と京都近郊に近づいた可視化結果 (2010)

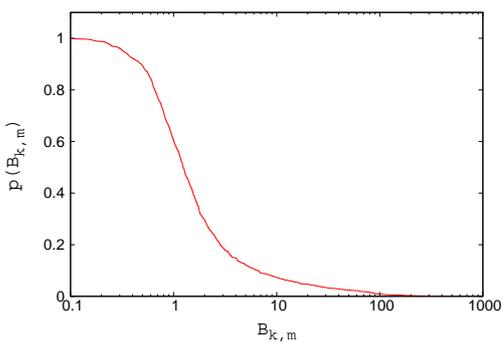


図 5:  $B_{k,m}$  以上のバースト度を持つ  $\mathcal{J}_{k,m}$  の存在確率

次に， $\mathcal{R}_k$  とバースト度  $B_{k,m}$  を伴う  $\mathcal{J}_{k,m}$  が提案可視化法によって，どのように視認でき，どのように視覚的な分析が行えるかを可視化結果を通じて検証する．

### 4.3 可視化結果

提案可視化法の有効性を検証するため， $K=205$  個の主要撮影スポット  $\mathcal{R}_k$  それぞれに対応する Glyph を 3D 空間上の日本地図上にマッピングした結果を図 6 に示す．提案可視化法では，全体的な傾向を見る場合ほど，可視化結果を鳥瞰視すれば良い．また，各地域の詳細や，各  $\mathcal{R}_k$ ，さらには  $\mathcal{J}_{k,m}$  を詳しく見たい場合ほど，注目する地域に近づけばよい．図 7 は，各地域に接近した例であり，Glyph がより詳細に視認できる．

鳥瞰視においては，旬シーズンの詳細に関する視認性が低くなるものの，年間に訪れる人が極めて多く， $V_k$  の大きい，つまりベースの高い  $\mathcal{R}_k$  が視認できる．さらに，シーズンが短期間であるほど  $B_{k,m}$  が大きくなる傾向を持つため，バースト性が極めて高い，つまりファイバーが長い  $\mathcal{J}_{k,m}$  は，ファイバーの色がほぼ単色になる傾向がある．そのため，鳥瞰視でも，長いファイバーは，容易に見え，色から，おおよその時期もわかる．ただし，ファイバーの色の視認性を高



図 8: 愛知県川見町の代表写真例 [2010]



図 9: 東京都立六義園の代表写真例 [2010]

めるため、鳥瞰視であるほど、太く見せている。

図 6 より、鳥瞰視において、東京や京都・大阪などの主要観光地に  $R_k$  が密集している様子が視認できる。また、長いファイバーを持つ  $R_k$  が主要観光地以外に散在し、近隣に撮影スポットが無く、孤立している場合もあり、それぞれファイバーの色が偏っていないことから、様々な時期で旬の度合いが高い  $J_{k,m}$  を持つ  $R_k$  が存在することも視認できる。つまり、各  $R_k$ 、個々を見ただけではわからず、周囲との空間的關係が見えてこそ気づく  $R_k$  の特徴を視認し得ることがわかる。

次に、複数の撮影スポットが密集する地域において、異なる撮影スポットの旬シーズンを比較することで、地域の全体的な傾向の違いを比較分析し得る事例を示す。そこで、東京と、京都に近づいた際の可視化結果を比較する。図 7(a) は、東京近郊、図 7(b) は、京都近郊に近づいたときの可視化結果を示している。図 7(a) と図 7(b) の比較から、年間の訪問者数が反映されたベースのみを比較すると、京都よりも東京の方が、ベースが極めて高い  $R_k$  が多いことがわかる。最も高いベースは浅草寺であり、2 位は渋谷であった。また、京都においては、最も高いベースが清水寺であり、清水寺のみが突出し、その他の  $R_k$  で  $V_k$  がそれほど多くないように見えることもわかる。次に、シーズンに着目すると、東京近郊に比べ、京都近郊の  $R_k$  は、 $V_k$  が相対的に低いものの、ファイバーは総じて東京近郊より高い傾向があり、突出したファイバーも多く、色から季節の違いもわかる。つまり、旬シーズンの傾向を比較することで、東京と京都の地域的な違いを読み取り得ることがわかる。

次に、図 7(a) と図 7(b) において、ベースが極めて高い撮影スポットは、ファイバーが低い傾向があるものの、よりファイバーに近づけば、いくつかのシーズン



図 10: 東京お台場の代表写真例 [2010]

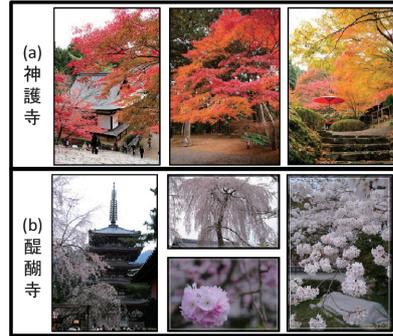


図 11: 京都神護寺と醍醐寺の代表写真の例 [2010]

があり、相対的にどのシーズンに人気があるかを知ることができることを示す。図 7(a) 中、(1) は浅草寺、(2) は渋谷のベース上部に近づいた例であるが、特に(2) の渋谷は、四つのシーズンがあり、それぞれファイバーの色が多色含まれる傾向があることから、長いシーズンが抽出されていることや、ファイバーの色から、おおよその時期も視認できる。さらに、各シーズンのパーセント性が顕著に違いを見せないことから、渋谷は、年中、同程度人が訪れている傾向があることがわかる。一方、図 7(b) 中、(1) は清水寺のベース上部に近づいた例であるが、六つのシーズンに分かれ、春や夏よりも、特に秋頃が最も顕著に人々を集めていることが視認できる。

以上から、1) 各スポットにいくつのシーズンが含まれるか、2) 各シーズンはどの時期か、3) 各シーズンはどの程度のパーセント性を示すか、4) 各スポットで、年間、どの程度人が訪れているか、5) スポット同士はどのような空間的關係にあるか、という五つの観点と同時に視認し得ることや、ベースやファイバーを比較することで、可視化結果から異なる地域の特徴の違いを分析できる点で、本研究の有効性が示唆された。

#### 4.4 議論

本研究の主題は、実世界の魅力や事象・事物を厳選して捉えるソーシャルセンサとしての撮影者たちについて、様々な行動を一望視できるよう可視化することにある。また、特徴的な撮影スポットや地域の発見、視覚的な分析、さらには特徴への気づきを仕向けることで、詳細な調査を促すことにある。提案法により、

この主題が果たされるかをより詳しく検証するため、バースト性の高いシーズンを持つ撮影スポットに着目して、写真の内容と時期・期間の情報から人手による調査を行い、旬シーズンの内容解釈を試みた。

日本全体を鳥瞰視した図 6 から、最もバースト度の高いシーズンを持つ撮影スポットは、図 6 中央付近にある、橙色の突出したファイバーであり、その季節が橙色であることから 11 月頃であること、また、この撮影スポットの周囲に他の撮影スポットが見えず、孤立していることが視認できる。ここで、孤立した地域に何があるのか、疑問が生じる。この撮影スポットは、調査をしたところ、愛知県の山間部、川見町周辺であり、交通便も悪いことがわかった。この撮影スポットに含まれる代表的な写真を図 8 に示す。図 8 には、11 月の山間部には珍しい紅葉が撮影されているが、よく見ると、満開の桜も同時に撮影されている。調査により、川見町には春だけでなく秋にも咲く珍しい四季桜が生息していることがわかった。また、この地域では、4 月にも桜が咲くものの、撮影行為は 11 月のみに行われていた。つまり、川見町は山間部の交通便の悪い地域であるにもかかわらず、この桜と紅葉が同時に観賞できる 11 月のみに人々を惹きつける特徴的な撮影スポットであることがわかった。

次に、図 7(a) の東京近郊において、緑のファイバーが突出している撮影スポットは、最も多い 14 シーズンが抽出された、都心近郊ながら自然にあふれた東京都立六義園であった。この六義園では、春の桜、ライトアップされた庭園の夜景、秋の紅葉など、数多くの旬シーズンを持つものの、図 9 のような緑が鮮やかな 7 月が最も旬の度合いが高いシーズンとして抽出された。また、図 7(a) の東京近郊において、黄色のファイバーが目立つ撮影スポットは、お台場であり、10 月頃に 2 つの突出した旬シーズンが存在した。最もファイバーが高かったのは、10 月 31 日にお台場で開催された、国内最大級のカーオーディオイベント Audio Car Gallery 2010 に対応しており (図 10(a))、2 番目にファイバーが高かったのは、10 月 2 日から 3 日に開催された日本最大級のモータースポーツに関するフェスティバル (モータースポーツジャパン) であり、都心近郊ながら、カーレースも開催される人気イベントであることがわかった (図 10(b))。これらは、必ずしも四季とは関係しない撮影スポットであると言える。

一方、図 7(b) の京都近郊においては、突出したファイバーが散見されるものの、上位 2 つを抜粋すると、一つは橙色のファイバーが突出した神護寺 (図 11(a)) であり、もう一つは紺色のファイバーが突出した醍醐寺 (図 11(b)) であった。神護寺は、京都における紅葉の名所として、ランキング 1 位になることもある名所中の名所であった。また、醍醐寺は、世界遺産であり、四季折々の風情が楽しめる撮影スポットであるものの、醍醐寺では、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 1 カ所で見ることができ、桜の名所中の

名所と言える。つまり、京都には、桜や紅葉の名所が多数存在するものの、突出したファイバーは、人々の実際の集団的行動が示している、際立った魅力的な旬シーズンを持つ撮影スポットの発見を促しており、提案可視化法の有効性が示唆される。

ただし、提案法で可視化された主要撮影スポットや旬シーズンは、Flickr データに基づいているため、人々の観光行動全般を必ずしも捉えきれていないかもしれない。しかし、近年、Geo-tag を付与できるカメラを搭載し、インターネットと常時接続できるスマートフォンが世界中で普及し、どこにいても写真共有サイトへ写真を投稿することが容易になってきている。また、Instagram など、新たな写真共有サービスも増え、さらには、お気に入りの場所 (Geo-tag) そのものを共有する Foursquare といったサービスも登場している。今後、Geo-tag は、人々の観光行動の傾向を反映する、より一般的な情報になると考えられる。したがって、本研究で提案した可視化法は、観光分析において、今後、重要性が増していくと考えている。また、本研究では、実験上、主要撮影スポットを抽出する際に、写真数 (人数に相当) に基づく閾値を設定している。一般的には、写真数が多いほど、ポピュラー性の信頼度は高まると考えられるが、閾値を上げるほど、極めて有名な撮影スポットばかりに注目し、隠れた名所を発見する機会を取り逃す場合も想定されるため、閾値の設定法や、埋没する名所を取り逃さない手法の探求が望まれる。また、本研究では日単位での旬シーズンを分析したが、「朝、昼、夜」等の時間帯に注目するものなど、様々な時間軸に基づく旬シーズンも考えられ得る。そのような観点からの旬シーズン可視化法を構築することも、今後の重要な課題である。さらに、近年、公共データの活用促進 (オープンデータ化) が推進されているが、それらを用いて提案可視化システムを拡張することにより、より高度化された観光支援法を探索することは、今後の重要な課題である。

## 5 まとめ

本研究では、多くの人々が実世界で撮影した、撮影位置、撮影時間 (年月日)、撮影者 ID に関するメタ情報が付与された大量の写真データから、人々の集団的行動を集約し、地域分析や観光への応用に向けて、時空間情報や撮影行動に関する多種情報の同時可視化問題に取り組んだ。そして、大量の写真データから、主要撮影スポットを抽出したあと、各主要撮影スポット独自のシーズンを抽出し、各シーズンの旬の度合いをバースト性の観点から定量化する手法を提案した。また、提案の有効性を検証するために、写真共有サイトから収集した大量の実データを用いて、提案法により抽出されたシーズンに関する評価を行い、提案可視化法による実験も行って、提案法の有効性を検証した。提案法により、発見や気づき、視覚的な分析を促せる事例を示し、地域分析や観光への応用に向けた有用性も示した。

ビッグデータ時代，インターネットには，より多  
 種で，無数のデータが蓄積され続けているため，応用を  
 検討し，さらなる効果や拡張性を求め，人々の行動情  
 報から新たな価値を創造する，より優れた手法を探索  
 する予定である．

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] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsu-  
 suo. Earthquake shakes twitter users: Real-time  
 event detection by social sensors. In *Proceedings  
 of the 18th International Conference on World Wide  
 Web*, pp. 851–860, 2010.
- [2] David J. Crandall, Lars Backstrom, Daniel Hutten-  
 locher, and Jon Kleinberg. Mapping the world’s  
 photos. In *Proceedings of the 18th International  
 Conference on World Wide Web*, pp. 761–770,  
 2009.
- [3] Slava Kisilevich, Florian Mansmann, and Daniel A.  
 Keim. P-DBSCAN: A density based clustering  
 algorithm for exploration and analysis of attrac-  
 tive areas using collections of geo-tagged photos.  
 In *1st International Conference on Computing for  
 Geospatial Research & Application*, 2010.
- [4] Slava Kisilevich, Florian Mansmann, Peter Bak,  
 Daniel A. Keim, and Alexander Tchaikin. Where  
 Would You Go on Your Next Vacation? - A Frame-  
 work for Visual Exploration of Attractive Places. In  
*GeoProcessing 2010*, pp. 21–26, February 2010.
- [5] 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.
- [6] Yuki Arase, Xing Xie, Takahiro Hara, and Shojiro  
 Nishio. Mining people’s trips from large scale geo-  
 tagged photos. In *Proceedings of the 18th Inter-  
 national Conference on Multimedia*, pp. 133–142,  
 2010.
- [7] Huagang Yin, Xin Lu, Changhu Wang, Nenghai  
 Yu, and Lei Zhang. Photo2trip: an interactive trip  
 planning system based on geo-tagged photos. In  
*Proceedings of the 18th International Conference  
 on Multimedia*, pp. 1579–1582, 2010.
- [8] Xin Lu, Changhu Wang, Jiang-Ming Yang, Yanwei  
 Pang, and Lei Zhang. Photo2trip: generating travel  
 routes from geo-tagged photos for trip planning. In  
*Proceedings of the 18th International Conference  
 on Multimedia*, pp. 143–152, 2010.
- [9] Zhijun Yin, Liangliang Cao, Jiawei Han, Chengxi-  
 ang Zhai, and Thomas Huang. Geographical topic  
 discovery and comparison. In *Proceedings of the  
 20th International Conference on World Wide Web*,  
 pp. 247–256, 2011.
- [10] Christopher G. Healey and Brent M. Dennis. Inter-  
 est driven navigation in visualization. *IEEE Trans-  
 actions on Visualization and Computer Graphics*,  
 Vol. 18, No. 10, pp. 1744–1756, 2012.
- [11] Stuart K. Card, Jock D. Mackinlay, and Ben Shnei-  
 derman, editors. *Readings in information visual-  
 ization: using vision to think*. Morgan Kaufmann  
 Publishers Inc., San Francisco, CA, USA, 1999.
- [12] Daniel A. Keim. Information visualization and vi-  
 sual data mining. *IEEE Transactions on Visualiza-  
 tion and Computer Graphics*, Vol. 8, No. 1, pp. 1–8,  
 January 2002.
- [13] Henrik R. Nagel, Erik Granum, Søren Bovbjerg,  
 and Michael Vittrup. Visual data mining. In  
 Simeon J. Simoff, Michael H. Böhlen, and Arturas  
 Mazeika, editors, *Visual Data Mining*, chapter Im-  
 mersive Visual Data Mining: The 3DVDM Ap-  
 proach, pp. 281–311. Springer-Verlag, Berlin, Hei-  
 delberg, 2008.
- [14] Dorin Comaniciu and Peter Meer. Mean shift: a ro-  
 bust approach toward feature space analysis. *IEEE  
 Transactions on Pattern Analysis and Machine In-  
 telligence*, Vol. 24, No. 5, pp. 603–619, 2002.
- [15] Dorin Comaniciu, Visvanathan Ramesh, and Peter  
 Meer. The variable bandwidth mean shift and data-  
 driven scale selection. In *Proc. 8th Intl. Conf. on  
 Computer Vision*, pp. 438–445, 2001.
- [16] David L. Davies and Donald W. Bouldin. A clus-  
 ter separation measure. *IEEE Trans. Pattern Anal.  
 Mach. Intell.*, Vol. 1, No. 2, pp. 224–227, February  
 1979.
- [17] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and  
 Luc Van Gool. Speeded-up robust features (surf).  
*Comput. Vis. Image Underst.*, Vol. 110, No. 3, pp.  
 346–359, June 2008.
- [18] 柳井啓司. 画像・映像の認識と意味的検索. *信学  
 技報 (SP2010-39)*, 第 110 巻, pp. 25–30, 7 2010.
- [19] Bernt Schiele and James L. Crowley. Object recog-  
 nition using multidimensional receptive field his-  
 tograms. In *Proceedings of the 4th European Con-  
 ference on Computer Vision - Volume I, ECCV ’96*,  
 pp. 610–619, London, UK, 1996. Springer-Verlag.

熊野 雅仁



1991年立命館大学工学部基礎工学科卒業。1991年龍谷大学工学部実験助手。2008年龍谷大学工学部実験講師。現在に至る。映像理論，コンテンツ制作支援技術，ソーシャル・メディア等に関する研究と教育に従事。博士(工学)[神戸大学大学院工学研究科情報知能学専攻]。情報処理学会，電子情報通信学会，日本音響学会，映像情報メディア学会，人工知能学会，認知科学会，芸術科学会，日本アニメーション学会，日本ヴァーチャルリアリティ学会，ACM(Association for Computing Machinery)，IEEE-Computer Science 各会員。

岩淵 聡



2012年龍谷大学工学部電子情報学科卒業。2014年龍谷大学大学院理工学研究科電子情報学専攻修士課程修了。

小関 基徳



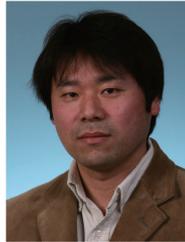
2012年龍谷大学工学部電子情報学科卒業。2014年龍谷大学大学院理工学研究科電子情報学専攻修士課程修了。

小野 景子



2003年同志社大学大学院工学研究科修士課程修了。2007年同志社大学大学院工学研究科博士課程修了。博士(工学)。2009年同志社大学研究開発推進機構省エネルギー照明システム研究センター特定任用研究員(助教)。2010年龍谷大学工学部電子情報学科助教。2014年龍谷大学工学部電子情報学科講師。並列処理，最適設計，進化計算等の研究に従事。電気学会，進化計算学会，建築学会各会員。

木村 昌弘



1987年大阪大学理学部数学科卒業。1989年同大学院理学研究科数学専攻修士課程修了。同年，日本電信電話株式会社入社。NTT コミュニケーション科学基礎研究所を経て，現在，龍谷大学工学部電子情報学科教授。複雑ネットワーク科学，データマイニングおよび機械学習の研究と教育に従事。博士(理学)。日本数学会，日本応用数理学会，人工知能学会，電子情報通信学会各会員。