

# 写真コミュニティサイトを使用した観光客数の推定方法について

日高亮太・磯田弦

## Estimation of visitors based on shared photos on the web

— a case of Kyoto using Flickr —

Ryota HIDAKA and Yuzuru ISODA

**Abstract:** This paper attempts to measure the spatial distribution of tourist in Kyoto (Japan), based on geographic coordinates of shared photos available on the internet. In order to remove the photos with incorrect coordinates, various filters were applied, and then verified with the tourist counts of the published tourism statistic. The approach had been able to identify popular places that are not listed in the statistic and those without recognizable place name.

**Keywords:** 写真コミュニティサイト (photo sharing web site), ジオタグ (geotag), 観光客 (visitors), 推定 (estimation)

### 1. はじめに

近年, 観光客の訪問先の把握のために GPS が活用されているが (藤田ほか, 2003; 野村・岸本, 2006), 個別の観光客に調査協力を募る方法はコストがかかり継続的な調査は困難である. 本研究では, Web 上で提供されている写真コミュニティサイトに投稿された写真の位置情報を GIS で使用し, 観光客数の推定を行う. 観光統計における訪問地調査では一般的に, 観光客に訪問地を記入してもらうアンケート調査が行われているが, 実際に観光客が訪問しているにも関わらず, 地名が知られていないまたは一義的に決まらないために, アンケートでは捕捉しきれない場所があると考えられる. 本論文では, 主に写真コミュニティサイトに投稿された写真の

中から, 観光客が撮影したと考えられる写真を抽出する方法を考察し, 空間的な観光客数の推計を行う.

### 2. 研究方法とデータ

Web に投稿された写真は, 観光客によって撮影されたものとは限らないし, 位置情報が記録されていてもそれが正しいとは限らない. そこで, 観光客の撮影した写真を適切に抽出する方法を考え, 既存の観光統計観光客数と比較することにより, 抽出方法の妥当性を検証する. そして最も適切な抽出方法で得た写真の情報を元に, 既存の統計書からは得られない側面での観光客数の推計を行う.

本研究で分析に使用する写真コミュニティサイトを Flickr とした. Flickr を使用する理由は, 写真の投稿枚数が世界最大であること, また Web API を広く公開しており, 写真を収集し易かったためである. Flickr の WebAPI を使用することで, 投稿さ

---

日高亮太 〒874-8577 大分県別府市十文字原 1-1

立命館アジア太平洋大学 磯田弦 宛

Phone: 0977-78-1111

E-mail: hidaka@tbz.t-com.ne.jp

れた写真のオーナー名，撮影日時，タイトル，座標値，タグなどを取得することができる。また，調査対象地は京都とした。理由は，写真の枚数が多いこと，日本の代表的な観光地であること，また観光統計が充実しているためである。

調査対象期間は2008年1月1日から2008年12月31までとした。そして，観光客推計の妥当性を検証するために，同期間を対象とした京都市産業観光局発行の平成20年度京都市観光調査年報（以下，調査年報）を使用した。

### 3. 写真データの性質

まず，投稿された写真から，月別，曜日別および時刻別の観光客数を検討する。分析には，位置情報が存在し「kyoto」のキーワードがある写真を対象とした。「kyoto」というキーワードを含む写真に限定した理由はこのキーワード無しの場合，京都居住者の写真が含まれる可能性が高く，また海外のWEBサイトのため，「京都」よりも「kyoto」のほうが写真数は多かったためである。

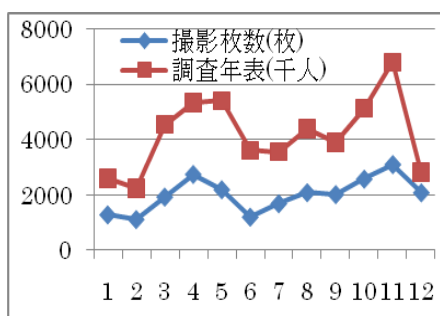


図-1 観光客数と撮影枚数

以上の条件を満たす京都市内で撮影された各月の写真枚数と調査年報に記載されている京都市への月間観光客数を比較したところ（図1）。この相関係数は0.87であった。写真の枚数から，京都への観光客数の推計が可能であると考えられる。

また，写真の撮影日時から，観光客の活動曜日（図

-2）や時間を推察することができる（図-3）。

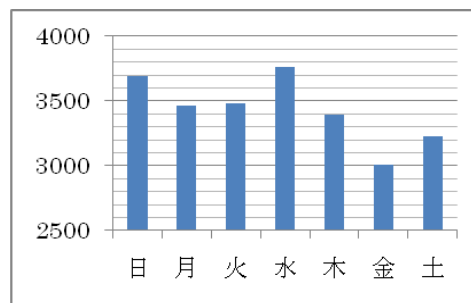


図-2 曜日別の写真枚数

図-3では，9時，13時，8時頃で写真の山があり，食事を撮影したのではないかと考えられる。また，深夜でも多くの写真が撮影されている理由は，外国人による投稿が多く，その外国人観光客のカメラの時刻設定に時差があるからだと考えられる。

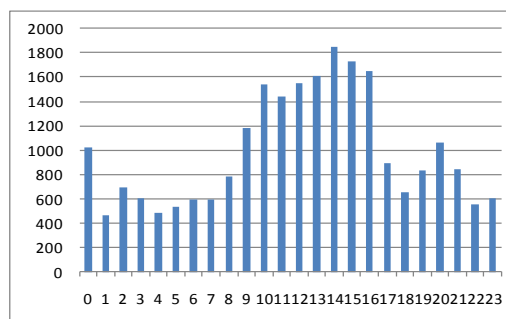


図-3 時刻別の写真枚数

写真の撮影日時と位置情報と組み合わせれば，原理的にはシーズン・月・時間ごとの観光客の空間的分布を把握できる。ただし，次節で述べるように，位置情報の正しい写真は十分に多くないため，現時点では残念ながら不可能である。

### 4. 写真の撮影地点と訪問観光客数との空間的分析

取得した写真の中には，位置情報が不正確な写真が含まれている。多くの場合，GPSを使用せずに，手作業で位置を確認し位置情報を付与された写真に位置情報が不正確なものが多い。そこで，分析に使用する写真に対しどのような条件を指定すれば位置情報の正確な写真だけを抽出することができる

かを検討する。適切な条件かどうかを検証するため、京都市観光調査年報掲載の観光客の訪問地上位 25 地点の訪問率と、25 地点での写真の撮影枚数を比べ、高い相関係数を示した抽出条件を適切だと判断する。

写真抽出の方法は、撮影日時、座標値、ユーザー名を用い、各種条件（フィルター）を作成し、その条件に当てはまるものを抽出するという方法を取った。表 1 にフィルターの条件とフィルターによって抽出された写真の枚数を記した。

表-1 フィルター条件

フィルター名 抽出枚数	フィルター条件 (分析に使用した写真)
フィルターなし 24042 枚	「kyoto」をキーワードに含み、位置情報（京都市内）がある写真
フィルター1 14899 枚	1 ユーザーにつき、1 日 1 枚の写真のみ使用
フィルター2 6050 枚	同一ユーザーの写真の座標値が重複している写真を削除
フィルター3 5196 枚	写真の座標値が重複している写真を削除（同一ユーザーに限らず）
フィルター4 6993 枚	異なる 2 地点以上で撮影しているユーザーの写真を使用

次に上位 25 観光地周辺の写真枚数を集計する。上位 25 観光地のポイントデータを作成し、そこを中心として半径 1000m, 500m, 250m, 100m, 50m のバッファを作成した。そしてバッファ上に何枚写真があるかを集計した。

表-2 は、調査年表の上位 25 観光地の訪問率と、同 25 地点バッファ内の写真の枚数の相関係数を各フィルターについてみたものである。これによれば、フィルター4 の 50m バッファの場合に最も高い相関係数が得られているが、写真枚数が少なくなってしまうため（表-3）、フィルター4 の 250m バッ

ファで集計するのが適当であると判断した。

表-2 フィルター毎の相関係数

	50m	100m	250m	500m	1000m
フィルターなし	0.729	0.539	0.355	0.232	0.026
フィルター1	0.657	0.493	0.300	0.192	0.011
フィルター2	0.646	0.631	0.625	0.533	0.353
フィルター3	0.645	0.635	0.644	0.591	0.460
フィルター4	0.731	0.654	0.673	0.626	0.477

表-3 フィルター4 の場合の使用バッファ半径ごと写真枚数（全地点合計）

検索半径	50m	100m	250m	500m	1000m
枚数	693	1429	2977	4111	5486

次に上位 25 観光地のバッファ内で撮影された写真枚数を 1 m<sup>2</sup>あたりの写真枚数密度に変換し、写真密度と調査年表に報告されている訪問率との回帰分析を行った。回帰分析では、貴船神社付近では写真が撮影されてなかったため、分析から除外した。また、写真密度と訪問率は両方とも、右に歪んだ分布であったため、自然対数をとって回帰分析に投入した（表-4、図-4）。

表 4 の回帰分析の結果を当てはめ、観光客訪問率の空間的推計を行った（図-5）。図-5 を確認すると、調査年報の上位 25 地点で高くなっているのはもちろんであるが、烏丸御池駅周辺の小路などの地名が知られていない場所や、四条通の河原町～八坂神社区間などの場所が地名によって一義的に定まらない場所でも高い値をとっている。これは、訪問地を問うアンケートからは集計できない、訪問先である。

## 5. おわりに

本研究では、写真投稿サービスの Flickr を利用し、観光客数の推計についての分析を行った。写真コミュニティサイトなどのユーザーが WEB 上に投稿し

た情報を副次的に利用し、分析に使用することは可能であると言える。しかし投稿された写真データには不正確なものも含まれており、それらをどう効果的に除去するかを検討した。この研究では、正確なデータを提供するユーザーを特定して、そのユーザーの写真データを使うのが有効であることがわかった。そして、既存の統計書には記載されていなかった、曜日、時刻や地図上での観光客数の推計を行うことができた。ただし、観光地の性質により、写真が撮られる枚数が異なるため、写真枚数を人数ベースに換算する際に誤差が発生するという問題点が残る。

表-4 訪問率の回帰：回帰係数（標準偏差）

切片	-1.452	(0.683) *
写真密度(自然対数)	0.184	(0.086) *
決定係数 R <sup>2</sup>	0.174	
予測の標準誤差	0.519	
F 値	4.630 *	
観測数	24	

\* 有意水準5%

従属変数: 訪問率(自然対数)

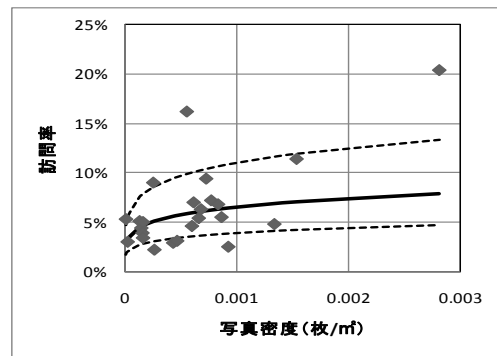


図-4 訪問率の予測値（破線は±標準誤差を示す）

#### 引用文献

野村幸子・岸本達也 2006, 「GPS・GIS を用いた鎌倉市における観光客の歩行行動調査とアクティビティの分析」総合論文誌 4, pp.72-77.

藤田朗・半明照三・山田雅夫・大内浩・三宅理一 2003, 「GPS 携帯電話を用いた回遊行動の調査分析」, 日本建築学会大会学術講演梗概集, F-1, 都市計画, 建築経済・住宅問題, pp.855-856

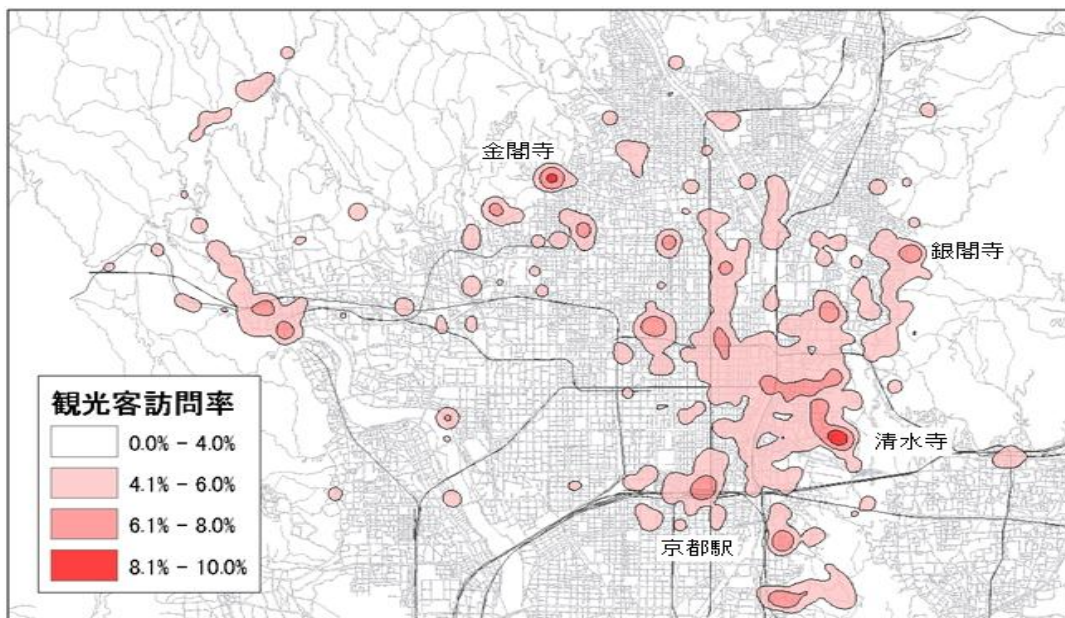


図-5 写真投稿枚数にもとづく観光客訪問率の推計値