

写真共有サイトに投稿された旅行写真群の時間信頼性向上

倉田陽平[†] 相尚寿[†] 真田風[†] 池田拓生[†]

首都大学東京 大学院都市環境科学研究科 観光科学域[†]

1. はじめに

旅行者の行動パターンや関心対象を知るための情報源として、ソーシャルビッグデータが注目を集めている。中でも Flickr 等の写真共有サイトには、世界各地のユーザが投稿した位置情報付き旅行写真が膨大に蓄積されているため、これをもとに、観光客や見所の分布推定、旅行者の移動軌跡の推定、旅程推薦ツールの開発などの取り組みが行われてきた[1]。このような活用例が有効に機能するためには、その前提として、元となる写真の時空間属性が精確でなければならない。しかし、時間属性に関しては、我々の標本調査において 2.66% の写真に昼夜逆転レベルの誤りが観察された (2 章)。この原因には、以下が考えられる ([2]を修正)。

- ▶ デジカメ内蔵時計に由来
 - ・ユーザによる時刻設定が不精確
 - ・タイムゾーン設定と現地時間の不一致
- ▶ GPS から得た時刻データに由来
 - ・撮影時刻ではなく、直近の GPS 信号の取得に成功した時点の時刻が記録されている
- ▶ そもそも撮影時刻ではない
 - ・スキャナーによる取込み時刻
 - ・Instagram 等のレタッチソフトによる保存時刻

管見の限り、これら誤差要因に対し積極的な対策を行った事例は見られない。すなわち、投稿写真を活用した既存の観光行動分析や観光情報ツールには、その信頼性に改善の余地が残されている。そこで本研究では、Flickr から抽出した日本国内の写真群の時間属性の信頼性を向上させる手法を開発する。ここで言う信頼性の向上とは、個別の写真について①タイムスタンプの修正が可能であるときは修正し、②修正が不可能なものはデータから除去することにより実現される。なお、処理高速化のため、個々の写真の Exif データおよび Flickr 側メタデータのみを利用し、画像処理的手法は使用しない。

2. 標本データ

Flickr API を利用し、2012 年 9 月または翌年 5 月に札幌・仙台・横浜・名古屋・京都・大阪・神戸・広島・福岡の各都心 3km 圏で撮影された写真と、それらを投稿した利用者 831 名が同日中に国内で撮影した位置情報付き投稿写真、合計 18,154 枚を抽出した。そしてこれらの写真を目視により、①明らかに昼の景色 (8206 枚) ②明らかに夜の景色 (866 枚) ③昼夜判別不能 (8082 枚) の 3 種類に選別した。そして、明らかに昼の景色なのにタイムスタンプが 20~3 時台のもの、ならびに夜の景色なのに 8~15 時台のものを「昼夜逆転写真」と判定したところ、483 枚 (2.66%) が該当した。

3. 三種類のタイムスタンプ

Flickr 上の写真は、そのメタデータに記録されたタイムスタンプのほか、Exif データ内に時刻情報を何種類か保有していることがある。このうち、撮影時刻の推測に有効だと思われるものとして、カメラ内蔵時計で計測された撮影時刻 *DateTimeOriginal* と、GPS センサにより計測された撮影時刻 (または直近の GPS 信号捕捉時刻) *GPSTimeStamp* に注目する。後者は通常、世界標準時で記録されているが、稀にローカル時刻となっていることがある。簡単のため、前者をカメラ時刻、後者に+9:00 を加え日本時間相当に変換したものを GPS_L 時刻と呼ぶ。図 1 に、今回の標本データ内写真の、カメラ時刻と GPS_L 時刻の保有状況を示す。

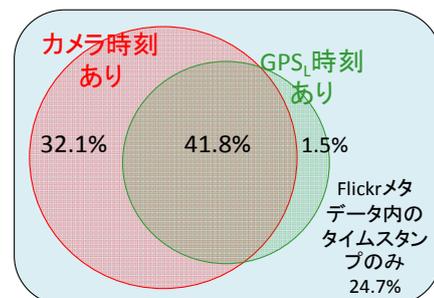


図 1. サンプル写真のタイムスタンプ保有状況

4. 提案アルゴリズム

以下の手順によって、撮影時刻の推定ならびに推定不可能な写真の排除を行う。参考のため、各過程で求められる時刻の観点から手順を整理したものを図2に示す。

Step 1. GPS_L時刻の修正

a) 同一カメラで撮影された一連の写真中にカメラ時刻・GPS_L時刻の両方を有する写真が複数ある場合、その差分の最小値・最大値を求め、その開きが60分以上であれば、一連の写真のGPS_L時刻を削除する

b) 上の差分が平均9時間程度であり、かつ2バイト文字を使用している者の写真は、GPS_L時刻を-9時間ずらす。

Step 2. カメラ・GPS_L時刻の統合

同一カメラで撮影された一連の写真中にカメラ時刻・修正GPS_L時刻の両方を有する写真がある場合は、その差分の最小値を求め、各写真のカメラ時刻に加算した「統合時刻」を求める

Step 3. タイムゾーン不明写真の除去

カメラ時刻しか持たない写真について、タイムゾーン設定が明らかに日本時間ではないと推測されるものを除去する

Step 4. 最終推定時刻の決定

各写真について、統合時刻、修正GPS_L時刻、選抜カメラ時刻*の各時刻が算出されていれば、この優先度順に最終推定時刻として採用する。何も採用できない場合は、その写真を除去する。
(*オプション)



図2. 提案アルゴリズムのフローチャート

Step 1-a では、信頼のおけない *GPSTimeStamp* を削除している。

Step 1-b は、GPS Time Stamp がローカル時として記録されるカメラを持つ日本周辺在住者への対処である。今回の標本データでは、GPS Time Stamp を持つ写真投稿者 253 名中 5 名が該当した。

Step 2 は、Thomee ら[2]の提案によるもので、GPS_L時刻の「概ね正確だが、測位不能な状況下

で遅れを生じたものが含まれるリスクがある」という性質と、カメラ時刻の「全体がずれている可能性があるが、相対的な時間関係は極めて正確である」という性質をかけあわせ、正確な撮影時刻を得ようとするものである。

Step 4 は、得られた各時間を正確な順から採用しようとするものである。ここでは統合時刻または修正GPS_L時刻の求まる写真だけでは枚数が不十分となる可能性を考慮し、オプションとして、カメラ時刻のみしか持たない写真について、明らかに日本時間では無いものを排除してから (Step 3) 追加採用を行えるようにした。

なお、Step 3 では「旅行者は深夜に撮影しない」という仮定のもと、同一カメラで撮られた一連の写真の中にカメラ時刻が深夜0~5時台のものが含まれれば、明らかに日本時間設定ではないと判定することにした。我々の標本データによれば、深夜0~5時台に撮影された写真は1.1%であり、仮定は妥当であると言える。

4. 評価

当初、18154枚の標本写真中、昼夜逆転写真は483枚(2.66%)含まれていた。提案アルゴリズム適用後(選抜カメラ時刻利用時)は、8,939枚(49%)が残存し、うち昼夜逆転写真は31枚(0.35%)まで減少した。一方、選抜カメラ時刻不使用の場合は、4,480枚(25%)が残存し、うち昼夜逆転写真は0枚となった。このことから、残存枚数より時間精度を重んじる場合は、選抜カメラ時刻を使用しないオプションが有効に働くことが確認された。

なお、Step 1-3において、修正GPS_L時刻、統合時刻、選抜カメラ時刻が求まったものはそれぞれ4,261枚、4,215枚、4,648枚で、それぞれの中の昼夜逆転写真は0枚、0枚、31枚であった。

5. おわりに

本研究の成果から、写真のメタデータのみをもとに、写真共有サイトFlickrから抽出した写真群の時間精度を向上できることが確認された。このアルゴリズムは、Flickrデータを活用した観光行動分析や観光情報ツールの信頼性向上への貢献が期待できる。

参考文献

- [1] 倉田陽平: 大量写真データをもとにした観光地内の主要観光ルート網の自動抽出に向けて. 観光情報学会第8回研究発表会, pp. 49-52, 2013.
- [2] Thomee, B., Moreno, J., & Shamma, D.: Who's Time Is It Anyway?: Investigating the Accuracy of Camera Timestamps. ACM Intl. Conf. on Multimedia, 2014.