

H-042

位置情報と大規模画像データベースを利用した撮影対象特定に関する検討 Object Identification using Location Information and Large-Scale Image Database

阿部 尚之[†]
Naoyuki Abe

大神 渉[†]
Wataru Oogami

島田 敬士[†]
Atsushi Shimada

長原 一[†]
Hajime Nagahara

谷口 倫一郎[†]
Rin-ichiro Taniguchi

1. はじめに

近年、世界中に存在するテキスト、画像、音声、映像など多様な情報に Web を通して容易にアクセスできるようになってきた。アクセス手段もオフィスなどのオフサイトでの利用を想定したデスクトップ型から、現場すなわちオンラインサイトで手軽に利用できるモバイル端末へと軸足移ってきている。それに伴って、コンテンツ管理者やユーザ自身が公開している静的な Web 情報のみではなく、より動的な即時性を要する実世界情報へのアクセス要求も高まっている。

我々は、実世界オブジェクトとのインタラクションを通して実世界情報を獲得する新たな枠組み「クリックブル・リアルワールド¹⁾ (以下、CRW と略記)」を提案している。これは、モバイル端末で実世界の対象を撮影する(シャッターを切る)ことにより対象を仮想的にクリックするというイメージを実現するものである。最近では、セカイカメラ²⁾のようにモバイル端末を通して、実世界空間とのインタラクション手段が提供されるようになりつつある。

CRW との相違点は、インタラクションの対象である。セカイカメラでは、空間に登録されたエアタグと呼ばれる情報にアクセスするため、空間とのインタラクションと言える。これに対して、CRW では、実世界中のランドマークなどの特定の対象物とのインタラクションを想定している。従って、ユーザは実世界対象物を撮影するだけで、その対象に関する情報を取得することが可能になる。システム側はシステム側が Web などからその対象に関する情報を収集するため、システム側に求められる課題は、ユーザが撮影した対象を特定して、その対象の情報を検索するためのキーワードを推定することである。

本稿では、ユーザの位置情報を効果的に利用して、大規模画像データベースに登録されている画像と画像に付加されたラベル集合から適切なキーワードを推定する方法を提案する。

2. クリックブル・リアルワールド

2.1 フレームワーク

CRW では、実世界中の対象を撮影することで対象に関する情報を検索することになる。すなわち、

1. モバイル端末で対象を「撮影する」というユーザのアクションを情報検索のトリガとする。
2. システムは撮影した画像に写った「対象」に関する検索キーワードを抽出する。
3. 検索キーワードを基に「対象」に関する情報を Web やセンシング Web から取得し、ユーザに提示する。というサービスが提供される。

2.2 システム側の処理の流れ

CRW では、ユーザのモバイル端末には GPS (位置情報システム) が搭載されており、対象を撮影する際に撮影位置情報も獲得できる環境を想定している。図 1 に CRW におけるシステム側の処理の流れを示す。

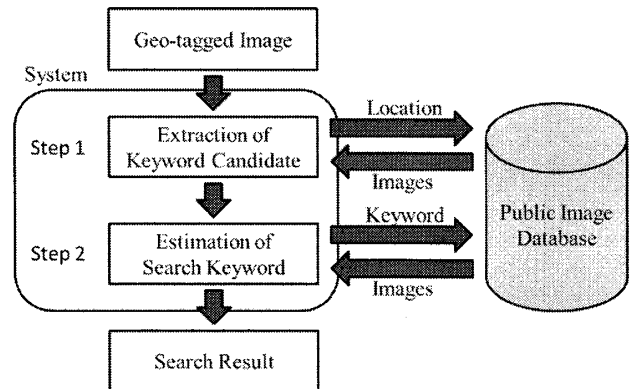


図 1 処理の流れ

まず、Step1 においてユーザが撮影した位置情報付き画像(ジオタグ画像)を取得し、位置情報を基に、Flickr³⁾などの公開画像データベースから、ユーザの位置情報をクエリとして、その周辺で撮影された画像を収集する。Flickr のような公開画像データベースを利用する利点は、各画像には撮影者が画像内に映っている対象を表すキーワードが付与されていることである。そのため、システム側で大量に画像を収集し、各画像を表すラベルを付ける必要がなくなる。Step1 では、位置情報を手がかりに収集した画像集合から、対象を表すキーワードの候補を複数抽出している。具体的には、収集された画像から抽出した画像特徴(SIFT特徴⁴⁾を利用)と画像に付与されているラベル特徴の共起関係を確率的正準相関分析⁵⁾により獲得している。得られた共起関係を基に、撮影画像の画像特徴に付与されやすいラベル特徴を推定することで、いくつかのキーワード候補を得ている。

Step2 では、キーワード候補群から、対象を最も良く表すキーワードを推定する。ここでは、位置情報を利用せずに、Step1 で抽出されたキーワードを画像検索のクエリとして利用している。キーワードをクエリとすることで、位置情報が付与されていない画像も含めて収集することができるため、Step1 で収集されるよりも多くの画像を収集することが可能である。従って、Step1 では位置情報を基にキーワード“候補”の絞り込みを行い、Step2 ではキーワード“候補”を利用した“キーワード”の絞り込みを行っている。具体的には、各キーワード候補で収集された画像から抽出した画像特徴(SIFT特徴)と撮影画像から抽出された画像特徴間の類似性を評価することで、最終的なキーワードの推定を行っている。

[†]九州大学 Kyushu University

2.3 これまでの問題点

これまでの手法⁹⁾では、図1のStep1において位置情報を基に Flickr から画像を収集する際に、ユーザの撮影位置を中心としてあらかじめ設定しておいた範囲内で撮影された画像を収集していた。実際に画像とそれに付与されたラベルを利用して確率的正準相関分析により画像特徴とラベル特徴の共起関係を学習し、撮影画像に付与されるべきラベルの推定確率を計算すると、その対象を表すラベルよりも、地名や撮影日時、exif 情報など、さまざまな画像に共通して付与されるラベルの推定確率が高くなってしまいう問題が生じていた。これは、地名や撮影日時、exif 情報がさまざまな画像に付与されることで、それらの出現頻度が高くなるのが原因である。

3. 生起確率の変化に着目したキーワード推定法

ユーザの撮影位置情報をクエリとした画像収集をする際に、これまでは固定したひとつの範囲に対して画像を収集していた。そのため、その範囲で撮影された写真に地名などのラベルが多数付与されていると、誤ったキーワード候補が抽出されてしまうことになる。そこで、本稿では、画像を収集する手法に改良を加える。具体的には、画像収集する範囲を段階的に広げることで、各範囲で収集された画像群から得られる各ラベルの生起確率の変化を調べる。例えば、「福岡」のような地名であれば、画像を収集する範囲を変えても、そのラベルを付与する可能性にはあまり変化が見られないことが期待できる。一方で、あるランドマークを表すラベルであれば、そのランドマークのごく近くではそれを表すラベルが付与されやすいが、遠くに離れるにつれて、そのようなラベルが付与される可能性は低くなると予想できる。

この考え方に基づいて、キーワード候補となるラベルの生起確率の変化を以下のように定式化する。撮影位置を中心として半径 R の範囲から収集した画像群を I_R とする。この画像群に付与されているラベル集合との共起関係を確率的正準相関分析により学習した結果を L_R とすると、撮影画像 T を入力とした場合にラベル l が生起する確率を $P(l|T, L_R)$ とする。ラベル l の生起確率 $P(l|T, L_R)$ が、画像収集半径を大きくするにつれて減少する度合いを、次の式で計算する。

$$D(l|T, L_i) = \frac{P(l|T, L_i) - P(l|T, L_{i+1})}{P(l|T, L_i)}$$

ここで、 i は段階的に変化させる画像収集半径であり、 i が大きくなるにつれて収集半径も大きくなる。 $D(l|T, L_R)$ の値が大きいほど、ラベル l の減衰傾向が大きいことを表すことになる。従って、画像を収集した半径すべてについて、 $D(l|T, L_R)$ を計算し、その重み付き総和

$$S(l) = \sum_i w_i D(l|T, L_i)$$

を各ラベルの減衰指数と定義する。また w_i は収集半径が大きくなるにつれて重みが小さくなるように $w_i = \exp(i/\sigma)$ とする。地名などのように比較的広範囲で共通して付与されるラベルは、この減衰指数が小さくなり、一方でランドマークのように特定の場所で付与される傾向にあるラベルについては、減衰指数が大きくなるが見込まれる。

4. 実験結果

広島県原爆ドームの写真のクエリとする実験を行った。画像を収集する半径 R を、 $R = \{100, 200, \dots, 700\}$ (単位はメートル)と100メートル刻みで7段階に変化させてラベルの生起確率の減衰指数を調べた。画像群に付与されているすべてのラベルの生起確率の減衰傾向を調べるのではなく、半径 $R = 100$ で収集した画像によって学習した結果 L_{100} を用いて、生起確率の高い10個のラベルについて減衰指数を調べた。実験では、 w_i を計算するときのパラメータ $\sigma = 200$ とした。

表1 従来法と提案手法によるラベル推定結果の比較

ラベル	半径100mでの生起確率 (従来法)	生起確率の減衰指数 (提案手法)
japan	0.899	0.0757
hiroshima	0.813	-0.0042
geotagged	0.223	0.0444
hiroshimaken	0.183	0.1360
peacepark	0.165	0.1682
原爆ドーム	0.132	0.2586
peacememorialpark	0.122	0.1911
asia	0.121	-0.0037
abombdome	0.119	-0.0907
日本	0.112	0.0579

表1に実験結果を示す。比較のため、従来手法でラベルを推定した結果を2列目に示している。従来法では、「japan」というラベルが最も高い確率として出力され、所望の「原爆ドーム」の生起確率は非常に小さな値になっている。これに対して、画像収集半径に応じたラベルの生起確率の減衰傾向を考慮した提案手法では、「原爆ドーム」の減衰指数が最も大きくなり、「japan」のような地名を示すラベルの減衰指数は小さくなっていることがわかる。このことから、ラベルの生起確率の減衰指数によりキーワード候補を精度良く絞り込むことが可能になることがわかる。

5. おわりに

本稿では、CRWにおける撮影対象の情報を取得するためのキーワード候補を効果的に抽出する方法について提案した。今後は、実験シーンを増やし、有用性の検証を進める予定である。

参考文献

- [1] 島田敬士, 大神渉, 谷口倫一郎, “クリッカブル・リアルワールド: モバイル端末を利用した実世界インタラクション”, 映像メディア処理シンポジウム(IMP2009) (2009).
- [2] “セカイカメラ”, available at <http://sckaicamera.com/>
- [3] “Flickr”, available at <http://www.flickr.com/>
- [4] D. G. Lowe, “Distinctive image features from scale invariant keypoints”, International Journal of Computer Vision, Vol.60 No.2, pp. 91-110 (2004).
- [5] 中山英樹, 原田達也, 國吉康夫, “大規模 Web 画像のための画像アノテーション・リトリバル手法”, 画像の認識・理解シンポジウム(MIRU2009) (2009).
- [6] 阿部 尚之, 大神 渉, 島田 敬士, 谷口 倫一郎: モバイル端末を利用した実世界インタラクションのための対象特定に関する検討, 信学技報パターン認識・メディア理解(PRMU2009-247), pp.85-90, (2010).