

クリッカブル・リアルワールド: 実世界情報獲得のための新たな実世界インタラクション

島田 敬士[†] 大神 渉^{††}
阿部 尚之^{††} 谷口 倫一郎[†]

本稿では、実世界情報を獲得するための新たな枠組み「クリッカブル・リアルワールド」を提案する。ユーザは携帯端末で実世界中の対象を撮影するというインタラクションを通してその対象に関する実世界情報を獲得することが可能になる。システム側に求められる技術課題には、撮影対象を認識し、その対象に関する検索キーワードを抽出することが挙げられる。本研究では、公開画像データベースを利用することでこれらの課題を解決する方法を提案する。

Clickable Real World: Acquisition of Real-world Information through Intuitive Interaction

ATSUSHI SHIMADA,[†] WATARU OOGAMI,^{††} NAOYUKI ABE^{††}
and RIN-ICHIRO TANIGUCHI[†]

We propose a new framework to interact with real-world objects. We call it “Clickable Real World”. A user can acquire information of a real-world object by taking an image with a cell phone’s camera. We regard the gesture of taking an image as clicking the real world. A system identifies the object based on image matching. We use open image database such as Flickr, Picasa, or so on, to identify the object. And then, we extract some keywords from the matched images. In this paper, we will explain the overview of Clickable Real World and report some experimental result.

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 を通じ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情報へのアクセス要求も高まっている。センシング Web^{1),2)} のようにセンサから得られる実世界情報を公開するという取り組みも盛んに行われていることから、今後はモバイル端末の携帯性を活かして、オンサイトで即時性の高い実世界情報に効率良くアクセスする仕組みが必要である。

そこで我々は、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「クリッカブル・リアルワールド」を提案する。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。ところで、モバイル端末上に実世界情報を AR 表示する先駆的な事例として暦本らによる NaviCam³⁾、最近では、セカイカメラ⁴⁾ が話題になっている。セカイカメラがインタラクションをしている対象はユーザが端末を利用している空間であり、空間中の明示的な対象ではない。そのため、ユーザが本当に獲得したい情報以外に多数のノイズ的な情報までもが提示されてしまうという問題がある。これに対して、クリッカブル・リアルワールドでは実世界空間中の具体的な「対象」まで特定したインタラクションを想定しているため、ユーザの意図をより反映した実世界情報獲得手段を提供することが可能となる。

[†] 九州大学 大学院システム情報科学研究院
Kyushu University

^{††} 九州大学 工学部 電気情報工学科
Kyushu University

2. クリックابل・リアルワールド

2.1 概要

従来の Web の情報検索では、検索フォームに検索したい情報のキーワードを入力することで検索を実行する方法が一般的である。一方で、クリックابل・リアルワールドでは、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。従って、ユーザが撮影した「対象」から、対象に関する情報を獲得するための「検索キーワード」を抽出することが本研究に求められる技術課題である。以下では、検索キーワードの抽出手法について述べる。

2.2 対象に関する検索キーワード抽出

Flickr⁵⁾ や Picasa⁶⁾ などの写真共有サイトで公開されている画像データベースを利用して、撮影対象を認識し、その検索キーワードを抽出する。これには次のような利点がある。

- 画像マッチングに必要な参照画像をシステム側が用意する必要が無い
- 公開画像には対象に関するキーワード(タグ)が付与されている

従って、公開画像データベースを利用することで、撮影対象の画像から検索キーワードへの変換を容易に実現することが可能になる。

しかしながら、公開画像データベースの画像をすべて利用して対象同定を行うことは、計算量の観点から現実的ではない。そこで本研究では、次のような流れで処理を行う。

位置情報を利用した対象同定

ユーザが対象を撮影した位置情報を利用して、公開画像データベースから参照画像を取得する際に、その位置情報付近で撮影された画像(以下ではジオタグ付き画像と呼ぶ)のみを取得する。これにより、画像マッチングの規模を実時間処理で実現可能な規模に縮約させることができる。撮影画像と高い類似度を持つジオタグ付き画像に付与されているタグ情報から対象に関するキーワード候補を抽出する。

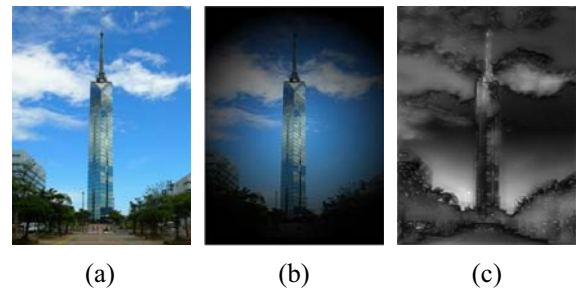


図 1 キーポイントの絞り込みのための画像処理：(a) 元画像，(b) 注視画像，(c) 顕著性画像

Fig. 1 Preprocessing for Refinement of Key Points: (a) Input Image, (b) Gazing Image, (c) Salient Image

キーワード候補を利用した対象同定

ジオタグ付きの画像は、ユーザが対象を撮影した位置周辺で撮影された画像であるため、撮影対象が含まれている可能性は高い。しかし、ジオタグ画像はジオタグが付与されていない画像(以下ではジオタグ無し画像と呼ぶ)に比べてその数が圧倒的に少ないというも現状である。そこで、上記で得られたキーワード候補を利用して、公開画像データベースからジオタグ無し画像を新たに参照画像として取得し、再度画像マッチングを行う。画像類似度の高い参照画像が多く得られれば、その画像を検索するために利用したキーワード候補は対象を表すキーワードである可能性が高いと言える。

このように、公開画像データベースの画像を、位置情報を利用して疎にマッチングし、その結果得られるキーワード候補を利用して再度画像を取得し、密にマッチングを行うというのが提案手法の特徴である。以下では、公開画像データベースを利用した対象同定(画像マッチング)とキーワード抽出の具体的手法について述べる。

2.2.1 対象同定手法

撮影画像と公開画像データベースから取得した画像から SIFT 特徴⁷⁾ を利用して画像間のマッチングを行う。具体的には SIFT アルゴリズムで検出した撮影画像と公開画像それぞれのキーポイントの特徴ヒストグラムを作成し、ヒストグラム間の Bhattacharyya 係数⁸⁾ により類似度を評価する。

また、計算量を削減し、マッチング精度を向上させるためには、特徴ヒストグラムを作成する前に不要なキーポイントを除外することが重要である。ユーザが撮影した画像と公開画像は「対象を中心」に「目立たせて」撮影していることが多いことを利用して、2種類の画像処理を施す。一方は、画像の中心から端に向

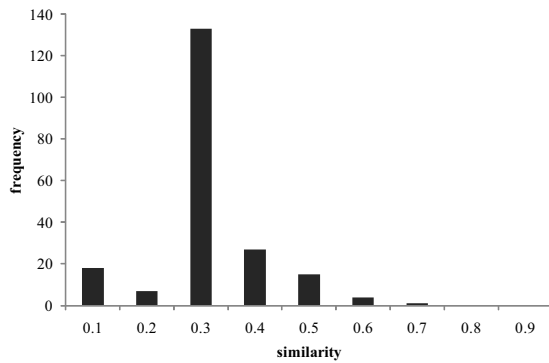


図 2 福岡タワージオタグ付き画像の類似度
Fig.2 Similarity of Geotagged Fukuoka Tower Images

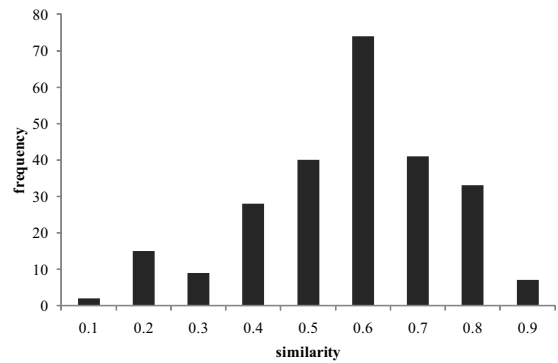


図 3 金閣寺ジオタグ付き画像の類似度
Fig.3 Similarity of Geotagged Kinkaku-ji Images

かうに連れて解像度を下げるガウシアンフィルタを施すことで画像の中心部分を注視した画像（注視画像）を生成する（図 1(b)）。他方は，周辺画素の輝度値間の顕著性を評価した画像を顕著性マップ⁹⁾に基づいて生成する（図 1(c)）。SIFT アルゴリズムで検出したキーポイントのうち注視度と顕著性の両方が高い画素のキーポイントのみを利用して特徴ヒストグラムを生成する。これにより，例えば，対象が中心に小さく写っている場合に，対象周辺のキーポイント特徴によって画像全体としての特徴が表現されてしまうのを防ぐことができる。

2.2.2 キーワード抽出方法

マッチした画像に付与されているタグの出現頻度の高い語をキーワードとして抽出する。ただし，地名やジオタグ特有のタグ情報等はあらかじめブラックリストに登録し，キーワードとして抽出されないようにしておく。

3. 実験

3.1 実験の設定

ユーザが撮影した画像として「福岡タワー」と「金閣寺」の 2 パターンを用意し，それぞれの写真が撮影された位置情報を利用して，Flickr からジオタグ付き画像を取得した。次に，画像マッチングにより撮影画像と画像類似度の高い画像のみを残して，タグの出現頻度を解析した。その後，出現頻度の高い単語（キーワード候補）を利用して Flickr から画像を取得し，再度画像マッチングを行った。

3.2 ジオタグ付き画像の類似度評価結果

福岡タワー 224 枚，金閣寺 250 枚のジオタグ付き画像を取得できた。これらのジオタグ付き画像と撮影対象画像との類似度を評価した結果（類似度の度数分布）を図 2 と図 3 に示す。グラフの横軸は類似度，縦軸は頻度を表している。金閣寺に関しては，福岡タワー

表 1 福岡タワージオタグ付き画像のタグ解析結果

Table 1 Tags of Geotagged Fukuoka Tower Image

ランキング	タグ	出現回数
1	geotagged	4
2	fukuoka	4
3	japan	3
4	panoramic	2
5	360 °	2
6	fukuokatower	2
7	hugin	2
8	d40x	2
9	geo:lat=335890767	2
10	福岡タワー	2
タグの総数		56
単語の種類		39

表 2 金閣寺ジオタグ付き画像のタグ解析結果

Table 2 Tags of Geotagged Kinkaku-ji Image

ランキング	タグ	出現回数
1	japan	39
2	kyoto	39
3	金閣寺	32
4	temple	29
5	kinkakuji	28
6	京都	14
7	日本	12
8	geotagged	9
9	goldenpavilion	8
10	rokuonji	7
タグの総数		380
単語の種類		123

よりも高い類似度を示すジオタグ付き画像が多く含まれていた。福岡タワーについては類似度 0.6 以上のジオタグ画像 5 枚，金閣寺については類似度 0.8 以上のジオタグ画像 41 枚を今回の実験ではマッチした画像と見なして以降の実験を行った。

3.3 キーワード候補の抽出結果

ジオタグ付き画像のうちマッチした画像に付与されているタグを解析した結果を表 1 と表 2 に示す。福岡タ

表 3 ジオタグ無し画像の類似度
Table 3 Similarity of Non Geotagged Images

キーワード	平均値	標準偏差	画像枚数
金閣寺	0.66	0.15	500
temple	0.48	0.17	500
goldenpavilion	0.58	0.15	499

ワーの場合, その固有名詞「福岡タワー」や「fukuoka-tower」の出現回数よりも「fukuoka」や「geotagged」などの地名やジオタグ特有のタグの方が出現回数が多かった。これはジオタグ付き画像の類似度評価結果で類似度が高い画像が5枚しか得られなかったことが原因の一つに挙げられる。ジオタグ付き画像の中には, 類似度が低くても福岡タワーを写した画像が他にも多数含まれていたため, 画像マッチング精度が向上させ, 検出もれを減らす必要がある。

一方で金閣寺の場合, ジオタグ付き画像で類似度が高い画像が41枚得られていた。地名のタグ「japan」や「kyoto」の方が「金閣寺」、「kinkakuji」、「goldenpavilion」よりも出現回数は高かったが, これらの地名タグはブラックリストに登録しているため, ユーザが撮影した対象に関するキーワード候補にはならない。

3.4 ジオタグ無し画像の類似度評価結果

先の実験で得られた金閣寺に関するキーワード候補のうち「金閣寺」、「temple」、「goldenpavilion」をFlickrの画像検索のキーワードとして利用して, ジオタグ無し画像を取得した。それらの画像とユーザが対象を撮影した画像との類似度評価結果を表3に示す。「金閣寺」あるいは「goldenpavilion」をキーワードとして取得した画像の方が, 「temple」をキーワードとして取得した画像よりも高い類似度を得ることができた。例として, 高い類似度が得られた画像を図4に示す。金閣寺が撮影されたジオタグ無し画像が正しくマッチングされていることが分かる。高い類似度を示したジオタグ無し画像を検索するために利用したキーワードは, ユーザが撮影した対象に関するキーワードである可能性が高いと判断することが可能である。以上のことから, ユーザが金閣寺を撮影した画像から最終的な検索キーワードを正しく抽出できることが確認できた。

4. おわりに

本稿では, 実世界中の対象を撮影するというインタラクシオンを通して実世界情報を獲得するクリックブル・リアルワールドの枠組みを提案し, その実現に向けた基礎的検討を行った。撮影対象を同定し対象に関するキーワードを抽出する実験では, 提案手法が有効に機



図 4 金閣寺ジオタグ無し画像でマッチした画像例
Fig. 4 Matched Images of Non Geotagged Images

能する場合とそうでない場合が結果として得られた。今後の課題としては, 実験シーンを増やすとともに, 画像マッチング精度の向上やキーワード抽出の精度向上が挙げられる。キーワード抽出に関しては, 単に出現回数を評価するだけでなく, 画像特徴とキーワードの出現パターンの相関などを考慮した評価方法を導入する必要がある。

参考文献

- 1) 美濃 導彦: “センシングウェブ: 概念と課題,” 人工知能学会誌, vol. 24, No. 2, pp. 179–184 (2009).
- 2) 谷口 倫一郎, 島田 敬士, 有田 大作: “センシングウェブにおけるセンサ情報の構造化,” 人工知能学会誌, vol. 24, No. 2, pp. 208–213 (2009).
- 3) Jun Rekimoto, “NaviCam: A Magnifying Glass Approach to Augmented Reality System,” Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp.399–412, MIT Press, 1997.
- 4) “セカイカメラ”, available at <http://sekaicamera.com/>
- 5) “Flickr”, available at <http://www.flickr.com/>
- 6) “Picasa”, available at <http://picasa.google.com/>
- 7) D. G. Lowe, “Distinctive image features from scale invariant keypoints”, International Journal of Computer Vision, Vol.60 No.2, pp. 91–110 (2004).
- 8) K. Fukunaga.: *Introduction to Statistical Pattern Recognition*, Academic Press, second edition (1990).
- 9) Laurent Itti, Christof Koch and Ernst Niebur. A Model of Saliency-based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp. 1254–1259 (1998).