

撮影位置情報と Web 検索を用いた写真及び映像インデキシング法

Indexing of Photo and Video Contents Using Shooting Position and Web Retrieval

岩崎 季世子† Kiyoko IWASAKI 山澤 一誠† Kazumasa YAMAZAWA 横矢 直和† Naokazu YOKOYA

1. はじめに

デジタルカメラや DV カメラ、カメラ付き携帯電話等の普及により、個人が写真や映像を撮影する機会は日常化してきている。その一方で、撮影した写真や映像を簡単に管理する方法は少なく、膨大な量のデータが未整理のままであることが多い。

写真や映像を簡単に管理する方法として、撮影した対象や時間、場所等についての情報をあらかじめ付加しておき、これに基づいた検索により所望の写真や映像を取得することが考えられる。JEITA 標準の Exif format[1] は、デジタル写真へのメタデータの記述について定めたもので、内容に関する記述や撮影時の焦点距離等のカメラパラメータ、GPS によって取得される位置情報等を画像ファイル自体に含めることができる。また、携帯電話で撮影した写真や映像に、GPS による位置情報を付加するという機能も普及してきており、今後、位置情報をもつ写真や映像が一般的になることが予想される。藤田らは、GPS やジャイロセンサ等の空間センサを用いて、高精度な空間情報付き写真を取得し、これを 3 次元実空間にマッピングすることで、位置情報に基づいた検索インタフェース [2] を提案している。

写真や映像に付加された位置情報は、地図データベースとのマッチングにより住所や施設名に変換できる。これにより利用者は、撮影位置に基づいた検索を容易に行うことができる。しかしこれは撮影した位置に基づくものであって被写体の位置に基づくものではない。このため、望遠や広角などで撮影を行った場合、被写体の説明には適さない情報が索引語として付加される可能性がある。また、地図データベースにない詳細な建物の情報や建物内にあるものについて説明する語を取得することはできず、地名や施設名だけでは、写真や映像の内容を説明する索引語として十分でない。このため、現在のところ内容に関するインデキシング作業は、人手によって行わざるを得ないことが多い。

本稿では、写真や映像の撮影位置・姿勢情報を利用して被写体の位置を推定し、推定位置の地名や施設名を用いて Web 検索を行うことにより、内容を説明する索引語を取得する手法を提案する。ユーザは、取得した索引候補語から、写真や映像に付加する索引語を選択する。これにより、ユーザのインデキシング作業を半自動化し、作業にかかる手間の軽減を目指す。

2. インデキシング手法の概要

図 1 に、撮影位置情報をもつ写真及び映像に対してインデキシングを行う手法の概要を示す。まず、GPS や

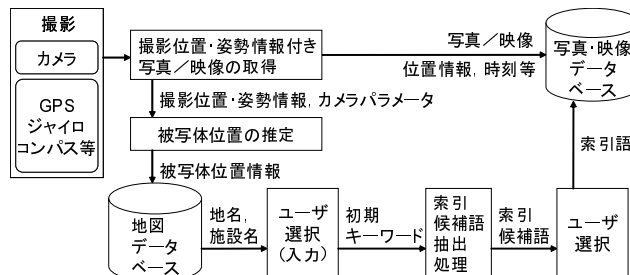


図 1: 撮影位置情報付き写真及び映像のインデキシング

ジャイロセンサ、コンパス等のセンサとカメラを用いて、位置・姿勢情報付きの写真や映像を取得する。ここで取得可能な撮影時刻、撮影地点の位置情報を写真や映像の索引とする一方で、この情報を利用して更に索引語を生成する。撮影地点の位置・姿勢情報とカメラパラメータから被写体の位置を推定し、地図データベースを参照することで、これを地名や施設名に変換する。ユーザは、取得した地名や施設名の中から 1 つを選択し、これを初期キーワードとして Web 検索を行い、検索された Web ページから索引候補語を抽出する。この候補語抽出処理については、3. で詳しく述べる。抽出された索引候補語から、ユーザは写真や映像に付加する索引語を選択する。これにより写真及び映像はその内容を記述する単語と関連付けられ、データベース内の写真や映像に対するキーワード検索が可能となる。

3. Web 検索による索引候補語の抽出

図 2 に、索引候補語を抽出するための処理の概要を示す。推定した被写体の位置情報は、地図データベースを参照して地名や施設名に変換する。この中からユーザが選択、または、直接入力した地名や施設名を初期キーワードとして Web 検索を行う。次に、検索されたページから HTML タグ等を削除してテキスト部分を抽出する。これに形態素解析を行って文を単語に分割し、各単語の品詞情報を得る。この品詞情報に基づき、索引語に適している名詞に分類された単語のみを抽出する。取得した全 Web ページにおける各単語の出現回数、出現ページ数を計測し、初期キーワードと抽出された各単語の各 Web ページにおける出現回数の相関係数を算出する。これらの情報から、以下の条件に合う単語を抽出する。

- (1) 複数の Web ページで出現する。
- (2) 初期キーワードと各単語のページ毎の出現回数の相関係数が、正の値をとる。

† 奈良先端科学技術大学院大学
Nara Institute of Science and Technology (NAIST)

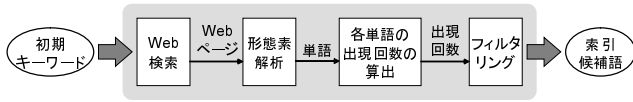


図 2: 索引候補語抽出処理

さらに、出現ページ数、相関係数、出現回数の順に優先的にソートを行う。この結果の上位単語をそれぞれキーワードとして Web 検索を行い、その検索件数を単語の特殊（一般）性の尺度として用い、検索件数の少ない、つまり特殊性の高い語を抽出する。これは、Web 上に頻繁に出現するような一般的な語ではなく、対象の写真や映像に特徴的な語を索引語として付加するためである。以上の処理により抽出した語を索引候補語としてユーザに提示する。

4. 実験

提案した手法のうち、被写体の位置情報に基づいた初期キーワードにより Web 検索を行い、索引候補語を取得する部分について試作システムを構築し実験を行った。試作システムに与える初期キーワードは、写真及び映像の被写体の位置が正確に推定できたものとして、市販の地図ソフト（アルプス社製「プロアトラス W2」）から被写体位置周辺の地名・施設名を距離の近い順に取得し、ここからユーザにより選択されたものとした。Web 検索には、Google の提供する Google API[3] を使用し、検索結果の上位 50 件のページを取得した。また、形態素解析には、日本語形態素解析システム「茶筌」[4] を使用した。

試作システムにより抽出された索引候補語を表 1, 2 に示す。表 1 は、初期キーワード「薬師寺」から索引候補語を抽出した結果である。「薬師寺」内の建物を示す「玄奘三蔵院」、「東塔」、「西塔」、建物を説明する「裳階」、「白鳳」、建物内にある「薬師如来」、「壁画」等の候補語が挙げられており、ユーザの索引語入力への補助となる単語が抽出できている。表 2 は、初期キーワード「奈良国立博物館」から索引候補語を抽出した結果である。「奈良国立博物館」内の施設を示す「ミュージアムショップ」、「新館」、「本館」等の候補語が挙げられているものの、展示品についての語は少なく、ユーザの入力補助として十分ではない。これは、形態素解析における辞書にそのような単語が登録されていなかったことが一因であると考えられる。

5. まとめ

本稿では、写真や映像の撮影位置・姿勢情報等を用いて被写体の位置を推定し、対応する地名や施設名を用いて Web 検索を行い、内容に関連する索引候補語を取得する手法を提案した。また、Web 検索を用いて索引候補語を取得する部分について試作システムを構築し、実験により写真や映像に対する索引語として有効な語が本手法によって抽出できることを確認した。

今後の課題としては、被写体の位置推定に必要なセンサとカメラからなる撮影システムの構築が挙げられる。

表 1: 索引候補語の抽出結果 「薬師寺」

索引候補語	検索件数	出現ページ数	相関係数	出現回数
裳階	504	8	0.386	14
玄奘三蔵院	896	8	0.378	26
東塔	2770	16	0.646	63
西塔	3640	15	0.645	41
中門	4440	7	0.547	15
天武天皇	7520	8	0.487	9
法相	7660	8	0.381	11
平山郁夫	8770	8	0.322	17
金堂	11100	14	0.707	51
薬師如来	13200	10	0.376	23
白鳳	14100	10	0.654	23
天平	18000	10	0.407	32
伽藍	18800	9	0.580	37
平城	19700	7	0.010	9
薬師寺	21200	45	1.000	242
皇后	33000	8	0.238	11
本尊	37200	12	0.489	22
壁画	38300	7	0.227	37
薬師	50900	12	0.388	20
講堂	55000	7	0.575	16

表 2: 索引候補語の抽出結果 「奈良国立博物館」

索引候補語	検索件数	出現ページ数	相関係数	出現回数
登大路町	3260	18	0.178	18
院展	3950	12	0.163	20
奈良国立博物館	5880	43	1.000	131
ミュージアムショップ	14000	7	0.061	9
水室	33800	9	0.032	23
新館	47500	8	0.081	15
考古	49800	6	0.473	10
東京国立博物館	59900	13	0.255	32
国宝	62800	9	0.094	15
倉	124000	12	0.163	20
仏教	132000	20	0.069	45
展覧会	132000	9	0.363	15
本館	155000	12	0.083	27
文化財	229000	18	0.259	33
祝日	264000	7	0.061	8
近鉄	324000	15	0.154	28
博物館	404000	37	0.113	270
大学生	420000	7	0.122	8
奈良	447000	34	0.266	180
地下	473000	7	0.015	11

また、Web 検索により取得した単語のデータベースを構築し、データベース内の各語に、索引語としての適性を示すパラメータを付加しておき、システム使用時に参照、更新を行うこと、また、ユーザによる索引候補語の選択を分析、学習することにより、抽出処理を効率化することや索引候補語の提示をより最適なものにすることを検討する。

参考文献

- [1] Japan Electronics and Information Technology Industries Association (JEITA): Exchangeable image file format for digital still cameras: Exif Version 2.2, 2002.
- [2] 藤田, 有川, 岡村: “高精度な空間情報付き写真の 3 次元実空間マッピング”, 電子情報通信学会論文誌 (A), Vol. J87-A, No.1, pp.120-131, 2004.
- [3] Google: Google Web API, <http://api.google.com/>
- [4] 松本: “形態素解析システム「茶筌」”, 情報処理, Vol.41, No.11, pp.1208-1214, 2000.