

撮影位置・姿勢情報に基づいた写真への索引付加システム

岩崎季世子[†] 山澤 一誠[†] 横矢 直和[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科

〒 630-0192 奈良県生駒市高山町 8916-5

E-mail: †{kiyoko-i,yamazawa,yokoya}@is.naist.jp

あらまし デジタルカメラをはじめとした撮影機器の普及により、個人が写真を撮影する機会は日常化してきている。しかし、写真を簡便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。写真を管理する方法の1つとして、写真の内容を説明する語を付加しておくことが考えられるが、これを人手で行うことは非常に手間がかかる。一方で、完全に自動化されたシステムによってユーザの意図した語を写真に付加することもまた、困難である。そこで本研究では、地図データベースと web 検索を用いた関連語抽出処理により撮影位置・姿勢情報に基づいた索引候補語を取得し、写真への索引付加作業を半自動化するシステムを提案する。索引付加作業は撮影と並行して行うことができ、撮影終了後すぐに効率的な閲覧が可能となる。また、プロトタイプシステムを用いて行った実験について報告する。

キーワード 位置・姿勢情報, 索引付加, 地図データベース, web 検索, 関連語抽出

An Indexing System for Photos Based on Shooting Position and Orientation

Kiyoko IWASAKI[†], Kazumasa YAMAZAWA[†], and Naokazu YOKOYA[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

Takayama 8916-5, Ikoma, Nara, 630-0192 Japan

E-mail: †{kiyoko-i,yamazawa,yokoya}@is.naist.jp

Abstract With the spread of digital cameras, shooting photos has been becoming an everyday affair. However, there are few methods or systems to manage photos simply, and a huge amount of photo data remains unorganized. Although it is possible to add appropriate words explaining the contents of the photo as one of the methods to manage photos, it requires much time and effort to input such indices manually. It is also difficult to add indices intended by a user automatically. In this paper, we propose a semi-automatic photo indexing system that enables the users to generate indices simply and browse a photo library efficiently. Index candidates are acquired by map database retrieval and relevant words extraction using web retrieval based on shooting position and orientation. The indexing work can be done in parallel with shooting, and efficient browsing is attained immediately after a shooting. We have implemented an indexing system based on the proposed method, and we report some experiments.

Key words shooting position and orientation, indexing, map database, web retrieval, relevant words extraction

1. はじめに

デジタルカメラやカメラ付き携帯電話等の普及により、個人が写真を撮影する機会は日常化してきている。その一方で、撮影した写真を簡便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。

画像の検索については、長年に亘ってさまざまな研究が行われている。画像の検索手法は、画像に含まれる色や形、テクス

チャといった画像特徴量を利用するものと、画像に付加された注釈や撮影場所、撮影時刻などのテキスト情報を利用するものに大別できる。前者には、ユーザが感性語を与えてそれに合う画像を検索するものや、与えられた画像に類似した特徴量をもつ画像を検索するものなどがある。後者は、ユーザが与えるキーワードでテキスト情報を検索し、これが付加された画像を取得するものである。

Mills ら [1] は上記の2つの手法を評価し、個人の写真に関し

ては、画像特徴量を使った検索よりも、画像にあらかじめ付加した注釈を利用したテキスト検索の方が、ユーザの意図した画像を取得できるとしている。

このような見地から、ユーザにとって負担の少ない形で写真に何らかのテキスト情報を付加し、効率的な管理を実現することを目的とした手法について研究が行われている [2] ~ [6] . Shneiderman ら [3] は、あらかじめデータベースに登録した人名等のラベルの一覧と、注釈付けを行う写真のサムネイル画像をユーザに提示し、注釈に適したラベルを画像上にマウスで drag-and-drop するシステムを提案している。また、自動的に注釈付けを行う方法として、Lieberman ら [5] は、撮影した写真をメールに添付して送るという行為を利用する手法を提案している。これは、写真を添付したメールの本文が写真を説明しているものとして出現する単語を写真の注釈とするものである。写真は、メールに添付して送られるごとにその内容を説明する文章が増えることとなり、より適切な検索が可能となる。しかし、Shneiderman らが提案するシステムは、注釈に使用するラベルをあらかじめデータベースに登録しておく必要があること、Lieberman らの手法は、メールに写真を添付するという限定された状況を想定しており、これ以外の状況では利用できないこと、全自動で処理されることによる注釈語の誤りなどの問題が挙げられる。

また、近年、撮影時の位置情報を利用して写真を管理する手法について研究が行われている [7] ~ [9] . JEITA 標準の Exif format [10] 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が一般的になることが予想される。写真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。これによりユーザは、撮影位置に基づいたキーワード検索や閲覧を行うことができる。しかし従来研究では、あらかじめ用意された地図データベースと写真の撮影位置をマッチングしているため、ユーザの意図した索引語が付加されない可能性がある。また、地図データベースにない単語については人手による入力が必要となる。

そこで本研究では、写真撮影時のカメラの位置・姿勢情報から被写体の位置を推定し、推定位置の地名や施設名の候補をユーザに提示することで、半自動的な索引付加作業を実現するシステムを提案する。索引候補語は、あらかじめ用意された地図データベースから対応する位置の地名や施設名を取得し、データベース内に適当な索引語が含まれていない場合には、web 検索を用いた関連語抽出処理によって新たな候補語を取得しユーザに提示する。ユーザに選択された索引語は、その位置に適当な語であると見なし、これを地図データベースへのフィードバックとして用い、データベースの更新を行う。これにより、提示される候補語が変化し、ユーザの選択作業が効率化されるものとする。

以降、2 章ではユーザが撮影した写真に対して、その位置・

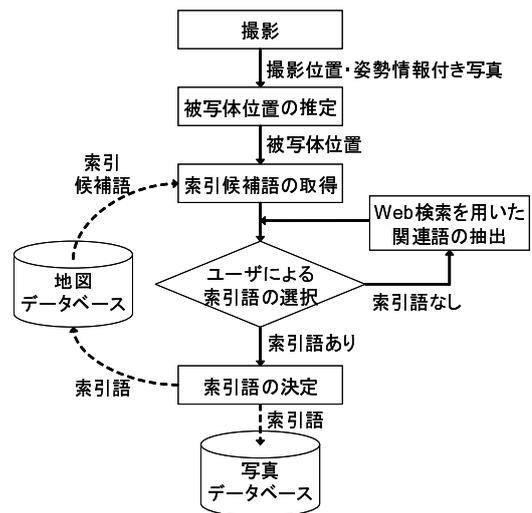


図 1 撮影位置情報付き写真の索引付加



図 2 地名・施設名の階層構造

姿勢情報に基づいて索引を半自動的に付加するためのシステムの構成と各部の特徴を説明する。3 章では、提案手法を用いたプロトタイプシステムの構築と実際に撮影した写真に索引付加を行った実験について述べる。

2. 撮影位置・姿勢情報に基づく索引付加

2.1 システムの概要

図 1 に撮影位置・姿勢情報をもつ写真に索引付けを行う手法の概要を示す。まず、GPS やジャイロセンサ、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を取得する。位置情報に関しては GPS を用いて、撮影位置の緯度・経度、標高を取得する。姿勢情報に関してはジャイロセンサ及びコンパスを用い、仰角、方位角を取得する。また、写真を記録した JPEG ファイルに含まれる Exif 情報から、カメラパラメータである焦点距離、レンズ F 値等を取得する。GPS、コンパス、ジャイロセンサは、カメラに固定し、撮影時の位置・姿勢情報を記録する。

撮影時に取得される撮影時刻、撮影地点の位置情報を写真の索引とする一方で、この情報を利用して索引候補語を取得する。写真の被写体について索引候補語を取得するため、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位置を推定する。この被写体位置の推定については 2.2 節で述べる。次に、推定した被写体位置を用いて地図データベースを参照し、推定位置付近の地名や施設名を取得する。地図データベースからの索引候補語の取得については 2.3 節で述べる。取得した地名や施設名は、写真の索引候補語としてユーザに提示され、ユーザは提示された候補語の中から被写体に適切な語を選択する。しかしこのとき、提示された候補語の中に写真に適した索引語が含まれていないということが想定される。これは、地図データベー

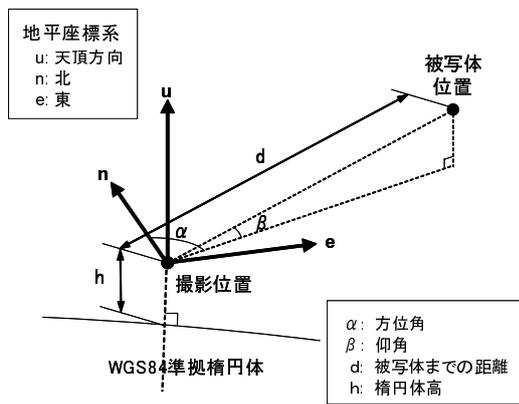


図 3 被写体位置の推定

すがその地点を代表する名称である地名や施設名を含むもので、施設内の建物といったより詳細なレベルの名称を含んでいないためである。例えば、図 2 に示すように、地図データベースは、上位層の「薬師寺」という施設名のデータを保持しているものの、下位層の名称である「薬師寺」内の建物の「金堂」、「東塔」、「西塔」といった施設名を保持していない。このような、より詳細なレベルの名称を新たに索引候補語として取得するため、web 検索を用いた関連語抽出を行う。ユーザは、提示された索引候補語の中から付加したいと考える索引語の上位層にあたる語を選択する。システムは、選択された上位層の語をキーワードとして web 検索を行い、関連語を抽出し、これを新たな索引候補語としてユーザに提示する。web 検索を用いた関連語抽出については 2.4 節で述べる。ユーザにより選択された索引語は、写真や撮影時刻・位置といったメタデータと共に写真データベースへ格納し、さらに、地図データベースへのフィードバックとしても利用する。選択された索引語を用いてデータベースの更新を行うことで、ユーザへの索引候補語の提示がより適切なものになっていくと考えられる。ユーザ選択の地図データベースへのフィードバックについては 2.5 節で述べる。

2.2 被写体位置の推定

図 3 は、撮影位置と被写体位置の関係を示したものである。まず、WGS84 (GPS の基準座標系) 上にある撮影位置の緯度・経度・標高を原点とした地平座標系を考える。地平座標系は、地表面付近のある点を原点として、天頂方向、東方向、北方向にそれぞれ軸をとる。この座標系上で、コンパス・ジャイロ・Exif 情報から得られる方位角・仰角・被写体までの距離を用いて推定被写体位置を算出する。これを WGS84 上に変換した緯度・経度を用いて以降の処理を行う。

2.3 地図データベースからの索引候補語の取得

前述の方法で推定した被写体位置を用いて地図データベースへの問い合わせを行い、写真に付加する索引の候補語を取得する。地図データベースは、初期状態では市販の地図ソフトに収録されている程度の地名や施設名から構成される。そして、索引付加システムがユーザに利用されることで、その入力内容をフィードバックとしてデータの追加・更新を行う。データベースの各レコードは、対象の名称とその緯度・経度、利用者によって選択された回数を用いた尤度からなる。これは、ユーザに選

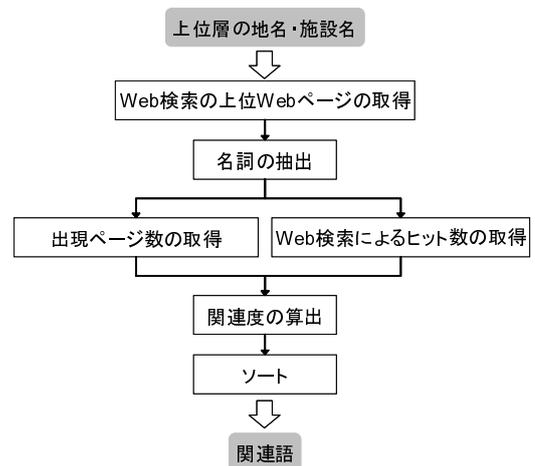


図 4 web 検索を用いた関連語抽出処理

択された回数が多いほど、その位置に尤もらしい索引語であるということを示す。推定した被写体位置とデータベースに収録されたデータの示す位置の間の距離を算出し、その距離の近い順に索引候補語としてユーザに提示する。

2.4 web 検索を用いた関連語の抽出

web を利用して関連語を自動収集することを目的として、佐藤ら [11] は、与えられた専門用語に対してその用語と関連する用語を web テキストを利用して収集する手法を提案している。佐藤らの手法では、収集した用語を辞典に利用することを目指し、正確な抽出を主眼としている。このため、処理にかかる時間は問題とされていない。一方、本研究で目的とするシステムでは、ユーザがインタラクティブに索引付加作業を行うため、処理にかかる時間を考慮する必要がある。このため本システムでは、抽出に使用する web ページ数を少なくし、また、web ページの解析による関連語の抽出は以下に述べる方法で行う。

図 4 に関連語を抽出するための処理の概要を示す。抽出処理の入力は、地図データベースを参照して取得した被写体位置付近の地名や施設名の中からユーザにより選択されたものである。まず、これを初期キーワードとして web 検索を行い、得られた URL の上位 10 件の web ページを取得する。次に、ページ内の HTML タグ等を除いた、タグ間のテキスト部分に対して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。これに基づき、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。ここで、初期キーワードと抽出した各単語の関連を示す指標として、式 (1) を考える。これは、web 内で初期キーワードの出現するページ集合と抽出した名詞 $word_i$ の出現するページ集合の積集合と和集合の比で表されており、この値が大きいほど初期キーワードと $word_i$ の関連は深いといえる。

$$\begin{aligned} relevance_i &= hit_{key \cap word_i} / hit_{key \cup word_i} \\ &= hit_{key \cap word_i} / (hit_{key} + hit_{word_i} - hit_{key \cap word_i}) \quad (1) \\ i &= 1, 2, \dots, N \end{aligned}$$

$relevance_i$: $word_i$ の関連度

hit_{key} : 初期キーワードの検索ヒット数

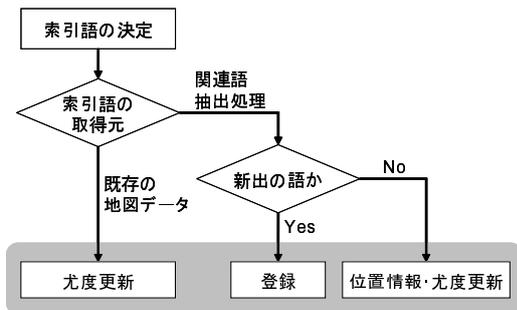


図5 ユーザ選択の地図データベースへのフィードバック

hit_{word_i} :抽出した名詞 $word_i$ の検索ヒット数

$hit_{key \cap word_i}$:初期キーワードと $word_i$ を含む検索ヒット数

$hit_{key \cup word_i}$:初期キーワードと $word_i$ のどちらかを含む検索ヒット数

N :抽出した名詞の数

算出には、新たに hit_{word_i} 、 $hit_{key \cap word_i}$ を取得する必要があるが、取得には非常に時間がかかるため、 $hit_{key \cap word_i}$ は式(2)より近似的に取得する。これは、 $hit_{key \cap word_i} / hit_{key}$ が、先に取得した初期キーワードを含む上位10件のwebページのうち $word_i$ が出現する確率とほぼ同じと仮定したものである。

$$hit_{key \cap word_i} \approx hit_{key} \times page_{word_i | key} / page_{key} \quad (2)$$

$page_{key}$:初期キーワードで取得したwebページ数
(現在は10)

$page_{word_i | key}$:初期キーワードで取得したwebページのうち $word_i$ が出現したwebページ数

以上により、 $page_{word_i}$ と hit_{word_i} を取得することで、関連度 $relevance_i$ を算出する。この関連度 $relevance_i$ で、抽出された名詞 $word_i$ をソートし、索引候補語としてユーザに提示する。

2.5 ユーザ選択の地図データベースへのフィードバック

図5は、ユーザが選択した索引語をフィードバックとした地図データベースの更新処理について示したものである。ユーザによって決定された索引語は、まず、その索引語の取得元によって処理を分岐する。索引語が市販の地図データや国土交通省提供の地図データなど既存の地図データから取得されたものである場合、その位置情報は信頼できると考えられる。そこで、索引語がユーザによって選択された回数を示す尤度を増加させる操作のみを行う。一方、索引語がweb検索を用いた関連語抽出処理により取得されたものである場合、その索引語が対応付けられる位置は写真の推定被写体位置である。このため、位置・姿勢情報の取得におけるセンサの精度によって、推定した被写体位置には誤差が生じている。これを考慮し、索引語が新出の単語である場合には対象とした写真の推定被写体位置をその単語の位置情報としてデータベースへの登録を行い、そうでない場合には式(3)により算出された値を新たな位置情報としてデータベースを更新する。また、尤度についても更新を行う。

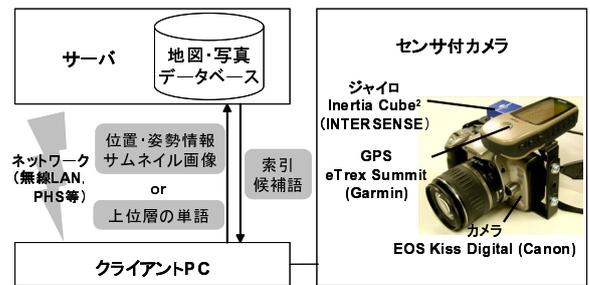


図6 プロトタイプシステムの構成



図7 索引語選択画面の外観

$$lat_{new} = (lat_{prev} \times lh_{prev} + lat_{sbj}) / lh_{new}$$

$$lon_{new} = (lon_{prev} \times lh_{prev} + lon_{sbj}) / lh_{new} \quad (3)$$

$$lh_{new} = lh_{prev} + 1$$

lat_{new}, lon_{new} :索引語に対応付けられる新たな緯度・経度

lat_{prev}, lon_{prev} :データベースに登録されている緯度・経度

lat_{sbj}, lon_{sbj} :被写体推定位置の緯度・経度

lh_{prev}, lh_{new} :尤度

以上のフィードバック処理が繰り返されることにより、新たに取得した索引語が対応付けられる位置は変化し、より適切な位置への修正が行われるものと考えられる。

3. 実験

3.1 プロトタイプシステム

提案したシステムのプロトタイプを作成し、撮影した写真への索引付加を行った。図6は、システムの構成を示したものである。カメラ、GPS、ジャイロからなるセンサ付カメラにより、写真と撮影時の位置・姿勢を取得し、クライアントPCに記録する。ユーザが、webブラウザを用いた入力画面で索引を付加する写真、位置・姿勢情報を記録したファイルを指定すると、これらの情報は、クライアントPCからネットワークを介してサーバに送信される。サーバはこれを処理して索引候補語を生成し、クライアントPCに送信する。

図7に、ユーザが索引付加作業を行うシステムの外観を示す。ユーザは、表示された写真に対して付加したい索引語を、プルダウン部分に提示される候補語から選択する。候補の中に付加したい索引語が含まれている場合には、それを選択して送信が



図 8 索引付加対象写真



図 9 索引候補語選択画面

タンを押すことで、写真データベースに写真と索引語、撮影時刻、撮影・推定被写体位置等の関連情報が格納される。候補の中に索引語が含まれていない場合には上位層の単語を選択して再取得のボタンを押すことで、web 検索を用いた関連語抽出処理によって得られる新たな候補語の取得をサーバ側に要求する。

地図データベースは、国土交通省の国土数値情報・施設データと市販の地図ソフト（アルプス社製「プロアトラス W2」）の施設データを使用し、ネットワーク上のサーバに置いた。web 検索を用いた関連語抽出処理には Google の提供する Google API [12] を検索エンジンとして使用し、検索結果の上位 10 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」[13] を使用した。

3.2 撮影した写真への索引付加

図 8 は索引付加の対象とした写真で、奈良の薬師寺において撮影したものである。用意した地図データベースには、索引候補語として「薬師寺」が含まれているが、薬師寺内にある個々

の建物の名称は含まれていない。以下では、図 8(a)～(d) の順に索引付加作業を行った様子について述べる。なお、今回の実験における索引付加作業は撮影終了後にあらためて行ったものである。

図 8(a) は「金堂」を撮影した写真であり、ここでは、この「金堂」という索引語を付加することを目指す。サーバは初期状態の地図データベースを保持しており、この写真を送信すると図 9(a) に示すように、「薬師寺」、「西の京派出所」などの索引候補語がユーザに提示される。ここでユーザが「金堂」の上位層の単語である「薬師寺」を選択し、再取得のボタンを押すと、サーバ側で「薬師寺」の関連語抽出処理が行われ、図 9(b) の画面がユーザに提示される。索引候補語の中には「金堂」が含まれており、ユーザがこれを選択し、送信することで写真データベースへの登録と地図データベースの更新が行われ、地図データベースには、表 1 の上段のレコードが追加される。また、こ

のとき「薬師寺」の関連語抽出処理により得られた候補語は、表 2 に示す単語を含む 402 語であり、その上位 100 語程度の中に薬師寺内のほとんどの建築物の名称が得られている。

次に図 8(b) は、図 8(a) と同様に「金堂」を撮影した写真で、図 8(a) とは別の方向から撮影されたものである。これをサーバに送信すると、図 9(c) の画面がユーザに提示される。地図データベースが図 8(a) 登録時に更新されたことから、1 度目の問い合わせから候補語として「金堂」が得られており、これを選択、送信するのみで写真データベースへの登録が完了する。また、ここでも地図データベースの更新が行われ、「金堂」に対応付けられる位置情報が、図 8(b) のもつ位置情報を利用して更新される。結果「金堂」のレコードは表 1 下段のように更新された。この更新前後の 2 点間の距離は約 45m である。なお、撮影時の GPS の誤差は約 10m であった。使用した 2 枚の写真の推定被写体位置間の距離は約 90m 離れており、推定精度が良いとはいえない。これは、算出に使用している Exif 情報から得られた被写体までの距離が、精度の高いものでないことによる。このため今後は、同じ対象を撮影した複数の写真の撮影位置・方向から、その交点を算出し、索引候補語の位置とする方法を検討したい。

同様の処理によって、図 8(c) については図 9(d) に示すように索引候補語「玄奘三蔵院」が選択され、索引が付加される。なお「玄奘三蔵院」は、索引候補語中の 45 番目に提示されている。一方、図 8(d) については図 9(e) に示すように、地図データベースへの問い合わせ、関連語抽出処理を行ったものの、適切な索引語である「興樂門」が得られなかった。これは、関連語抽出処理において取得した上位 10 件の web ページに「興樂門」という単語が含まれていなかったことによるものである。

表 1 地図データベースの更新

入力	name	north_south	latitude	east_west	longitude
図 8(a)	金堂	N	34 °40' 4.8"	E	135 °47' 3.4"
図 8(b)	金堂	N	34 °40' 6.0"	E	135 °47' 4.2"

表 2 抽出された「薬師寺」の関連語 (上位 20 語)

1	坐像	11	天平
2	東院堂	12	尊像
3	フェノロサ	13	西ノ京
4	薬師如来	14	天武天皇
5	光明皇后	15	修二会
6	三重塔	16	奈良駅
7	東塔	17	金堂
8	西塔	18	白鳳
9	吉祥天	19	催事名
10	会式	20	立像

4. ま と め

本稿では、写真の撮影位置・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の候補をユーザに提示することで、半自動的な索引付加作業を実現するシステムを提案

した。索引候補語は、あらかじめ用意された地図データベースから対応する位置の地名や施設名を取得し、写真に適当な索引語が含まれていない場合には、web 検索を用いた関連語抽出処理によって取得した。また、ユーザに選択された索引語は、その位置に適当な語であると見なし、これを地図データベースへフィードバックさせることで提示される候補語を変化させ、ユーザの選択作業の効率化を図った。

提案したシステムのプロトタイプを用いて、撮影した写真に対して位置・姿勢情報に基づいた索引語の付加を行い、適切な索引語の付加が可能であることを確認した。また、システムの利用により、地図データベースへの新たな地図データの追加や更新が行われることで、索引候補語をより適切に提示することができたと考えられる。

今後の課題としては、まず、地図データベースと被写体位置のマッチング方法の検討が挙げられる。現在は、推定した被写体位置から一定距離内の候補語を近い順に取得しているが、今後は、撮影方向や撮影時のセンサの受信状況による誤差や被写界深度、被写体の大きさ等を考慮した方法を検討する。また、ユーザの利用回数を示す尤度を用いた提示順序の決定方法を検討する。さらに、場所・利用者に関してより広範な評価実験を行う必要がある。

文 献

- [1] T. J. Mills, D. Pye, D. Sinclair and K. R. Wood: "Shoebbox: A digital photo management system," Technical Report 10, AT&T Laboratories Cambridge, 2000.
- [2] A. Pentland, R. W. Picard and S. Sclaroff: "Photobook: Content-based manipulation of image databases," Int. Journal of Computer Vision, Vol. 18, No. 3, pp. 233-254, 1996.
- [3] B. Shneiderman and H. Kang: "Direct annotation: A drag-and-drop strategy for labeling photos," Proc. of the Int. Conf. on Information Visualization, pp. 88-95, 2000.
- [4] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski and B. Field: "Semi-automatic image annotation," Proc. of the Human-Computer Interaction - Interact'01, pp. 326-333, 2001.
- [5] H. Lieberman, E. Rosenzweig and P. Singh: "Aria: An agent for annotating and retrieving images," IEEE Computer, Vol. 34, No. 7, pp. 57-61, 2001.
- [6] D. Frohlich, A. Kuchinsky, C. Pering, A. Don and S. Ariss: "Requirements for photware," Proc. of the 2002 ACM Conf. on Computer Supported Cooperative Work, pp. 166-175, 2002.
- [7] K. Toyama, R. Logan, A. Roseway and P. Anandan: "Geographic location tags on digital images," Proc. of the eleventh ACM Int. Conf. on Multimedia, pp. 156-166, 2003.
- [8] D. D. Spinellis: "Position-annotated photographs: A geotemporal web," IEEE Pervasive Computing, Vol. 2, No. 2, pp. 72-79, 2003.
- [9] M. Naaman, Y. J. Song, A. Paepcke and H. Garcia-Molina: "Automatic organization for digital photographs with geographic coordinates," Proc. of the 2004 Joint ACM/IEEE Conf. on Digital Libraries, pp. 53-62, 2004.
- [10] J. Electronics and I. T. I. Association(JEITA): "Exchangeable image file format for digital still cameras: Exif version 2.2," 2002.
- [11] 佐藤, 佐々木: "ウェブを利用した関連用語の自動収集", 自然言語処理, Vol. 153, No. 8, pp. 57-64, 2003.
- [12] Google: Google Web API, <http://api.google.com/>.
- [13] 松本: "形態素解析システム「茶釜」", 情報処理, Vol. 41, No. 11, pp. 1208-1214, 2000.