

撮影位置・姿勢情報に基づく写真への索引付加

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。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 撮影位置・姿勢情報, 写真への索引付加, 地図データベース, web 検索, 関連語抽出

Indexing Photos Based on Shooting Position and Orientation

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

[†] 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 indexes manually. It is also difficult to add indexes intended by a user automatically. In this paper, we propose a semi-automatic photo indexing system that enables users to generate indexes simply and browse a photo library efficiently. The proposed system is based on geographical database and web retrieval using shooting position and orientation information.

Key words shooting position and orientation, photo indexing, geographical database, web retrieval, relevant word extraction

1. はじめに

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。

画像検索については、長年に亘って研究が行われており、さまざまな手法やシステムが提案されている [1, 2]。また、web 上で画像の検索を行う検索エンジン [3, 4] や、画像 (写真) 閲覧用のソフトウェア [5, 6] も一般に普及している。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によるメタデータを利用するものに大別できる。前者には、ユーザが描いたスケッチや与え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 [7-10]。後者は、ユーザが与えたキーワードで、画像に付加されたテキスト情報を検索し、該当する画像を取得するものである [11, 12]。

写真についても、これらの手法を適用した検索や閲覧を行うことができるが、本研究では個人が撮影した写真を対象としており、これに適切な手法を検討する。個人が撮影する場合の被写体としては、家族や友人などの人物、旅行先などの風景や施設等が一般に想定される。また、撮影した写真を閲覧・検索する状況としては、撮影した本人やその身近な人物が、特定の日時・人物・出来事・場所などに基づいて行うことが想定される。

撮影した日時については、写真のメタデータとして付加されている場合がほとんどであり、日時に基づいた閲覧・検索は一般に行われている。人物については、個人の写真であれば、家族や友人などある程度限られた人物のみが撮影されているものと考えられる。そこで、顔認識手法を用いたシステム [13] や写真の人物にラベル付けを行うインターフェース [14] が提案され、人物に基づいた閲覧・検索を実現している。出来事については、個人のスケジュール管理を行うソフトウェアを参照して、撮影日時に対応する予定から写真の場面を類推する手法 [15] が提案されているが、適用できる状況は限定されている。場所については、地図を用いた GUI から写真の撮影位置をユーザが手入力するシステム [16] や GPS を利用して位置情報を取得するシステム [15, 16] などが提案されている。このようなシステムでは位置情報が数値として得られるが、閲覧や検索におけるユーザの入力は、一般に地名や施設名である。したがって、得られた位置情報を閲覧や検索に利用するためには、位置情報を数値から地名や施設名といったテキスト情報に変換する必要がある。この変換には地図データを利用することが考えられるが、対応するデータがない場合やユーザの意図とは異なるデータに対応付けられる場合など、適切な変換が行われないことがあり、場所に基づいた閲覧や検索を行うには課題がある。

一方で近年、写真に対する撮影位置情報の付加は一般的になりつつある。例えば、市販されているデジタルカメラにおけるメタデータの標準規格である Exif [17] は、カメラパラメータや GPS で取得した位置情報、写真の内容に関する記述などを JPEG 形式の画像ファイル自体に含めることができる。また、GPS とカメラの付いた携帯電話により位置情報付きの写真を取得することができる。そこで本研究では写真撮影時のカメラの位置・姿勢情報に着目し、この情報を用いて被写体位置を推定し、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切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、その位置に適切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使い、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、ユーザの選択作業は効率化される。

以降、2 章では撮影位置・姿勢情報に基づく写真への索引付加手法について、3 章では提案するシステムのプロトタイプを用いた実験について述べる。最後に 4 章で本稿をまとめ、今後の展望について述べる。

2. 撮影位置・姿勢情報に基づく索引付加

2.1 索引付加の概要

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。はじめにユーザは、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撮影位置・姿勢情報付きの写真を取得する。撮影時に取得される撮影時刻、撮影地点の位

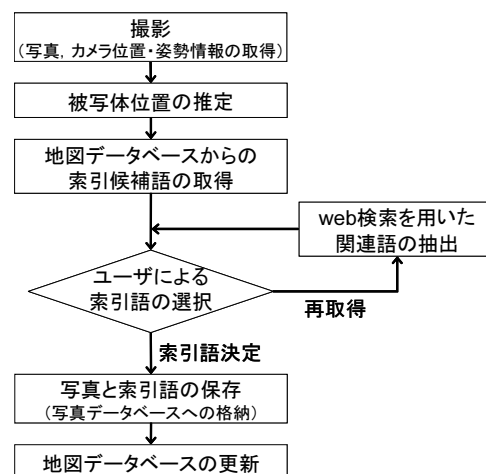


図 1 撮影位置情報付き写真への索引付加

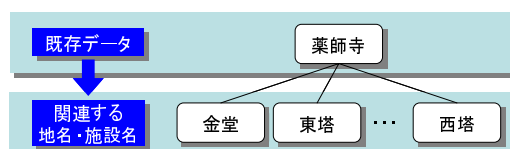


図 2 関連する地名・施設名の取得

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。

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。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、地図データベース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「薬師寺」という施設名のデータを保持しているものの「薬師寺」内の建物である「金堂」「東塔」「西塔」といった施設名を保持していない。そこで、より詳細なレベルの名称をはじめとした、関連する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に最も関連すると考えられる語を選択する。システムは、選択された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。ユーザにより選択された索引語は、写真や撮影時刻・位置情報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、ユーザに提示される索引候補語やその提示順序を変化させ、ユーザによる作業の効率化を図る。上記の手順で、写真とその索引語およびその他のメタデータの格納された写真データベースが構築され、ユーザ

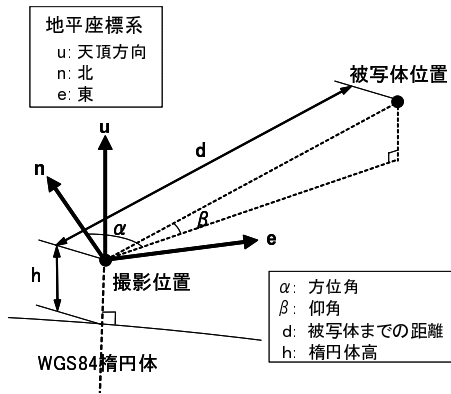


図 3 被写体位置の推定

は、索引語やメタデータに基づく閲覧や検索を行うことができる。

2.2 撮影位置・姿勢情報付き写真の取得

提案システムを利用するユーザは、撮影位置・姿勢情報を取得することのできるセンサ付のカメラを用いて撮影を行う。本研究では、カメラに GPS やジャイロセンサ、コンパス等が取り付けられていることを想定する。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およびコンパスを用い、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。

2.3 被写体位置の推定

撮影時に取得した位置・姿勢情報から、撮影された被写体の位置を推定する。図 3 に、撮影位置と被写体位置の関係を示す。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、2.4 節で述べる地図データベースからの索引候補語の取得、および 2.6 節で述べる地図データベースの更新を行う。

2.4 地図データベースからの索引候補語の取得

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。

2.4.1 地図データベースの構成

地図データベースの各レコードは、索引候補語となる地名や施設名といった名称、対応する位置の緯度・経度、ユーザに選択された回数を表す頻度から構成される。表 1 に例を示す。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たデータなどの既存のデータを格納したものであるが、索引付加システムをユーザが利用することで、新たなデータの登録や既存のデータの更新が行われる。データの登録や更新は、ユーザが索引付加作業において選択した索引語に基づいて行われる。詳細については 2.6 節で述べる。

表 1 地図データベースに含まれるデータの例

name	lat[deg]	lon[deg]	freq
薬師寺	34.668878	135.784313	0
東塔	34.668073	135.784335	5
唐招提寺	34.675775	135.784786	4
五重塔	34.614133	135.733928	13

2.4.2 位置情報と選択頻度に基づく索引候補語の取得

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。まず、推定した被写体位置からの距離が一定範囲内の地名や施設名を近い順に取得する。次に、撮影位置と候補位置を結ぶベクトルと撮影時の姿勢情報として取得した撮影ベクトルの方位角成分について類似度を算出する。推定被写体位置からの距離と方位角の類似度を考慮し、式 (1) により算出される値により候補を並べ替える。式 (1) は 0 から 1 の値をとり、値が大きいほど索引付加対象とする写真に適するものとする。

$$likelihood_i = 0.5 \times (1.0 - distance_i / \max(distance)) + 0.5 \times |\cos\theta_i| \quad (1)$$

$(i = 1, 2, \dots, N)$

但し、

$likelihood_i$: 候補の写真への適切さを示す評価値、

$distance_i$: 推定被写体位置と候補位置間の距離、

θ_i : 撮影位置と候補位置を結ぶベクトルと

撮影ベクトルの方位角成分のなす角

N : 索引候補語の数。

さらに、取得した候補をユーザによる選択の有無で分類し、これまでにユーザに選択された履歴のある候補を上位に、選択された履歴のない候補を下位にし、索引候補語としてユーザに提示する。

2.5 web 検索を用いた関連語の抽出

提案するシステムでは、写真に対して撮影位置・姿勢情報に基づく索引語を付加することを目的とするが、地図データベースに該当する位置のデータがない場合や取得した索引候補語が写真に適切でない場合には、新たな候補語を何らかの手段で取得する必要がある。本研究ではその手段として web 検索を用いた関連語の抽出を検討する。web 検索では、ある程度の知名度をもつ対象であれば詳細な情報を得ることができる。本研究で着目する被写体に対応する地名・施設名についても、地図データより詳細な情報を web 上から取得することが期待される。

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 [18] は、与えられた専門用語につい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旨とし、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

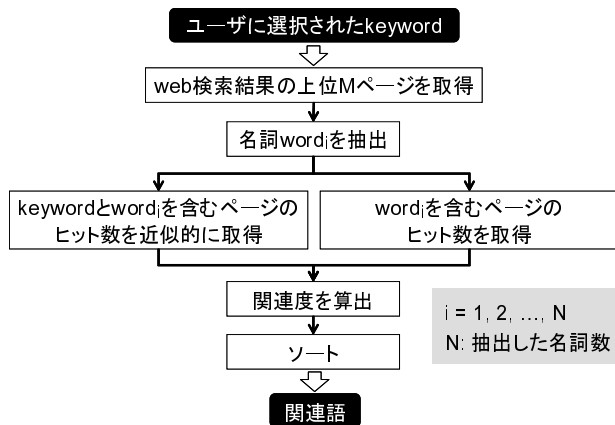


図4 web 検索を用いた関連語抽出処理

ムでは、web 検索の結果として取得される web ページのうち、抽出に使用するページ数を少なくし、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に関連語抽出処理の概要を示す。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ーにより選択された単語である。この選択された単語を以下「キーワード」と呼ぶ。まず、このキーワードを用いて web 検索を行い、得られた URL の上位 M 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、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(2)で定義される関連度を考える。これは、web 内で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キーワードと $word_i$ の関連は深い。

$$rel_i = \frac{hit_{key \cap word_i}}{hit_{key \cup word_i}} = \frac{hit_{key \cap word_i}}{(hit_{key} + hit_{word_i} - hit_{key \cap word_i})} \quad (2)$$

$(i = 1, 2, \dots, N)$,

但し、

rel_i : $word_i$ の関連度,

hit_{key} : キーワードを含むページの検索ヒット数,

hit_{word_i} : $word_i$ を含むページの検索ヒット数,

$hit_{key \cap word_i}$: キーワードと $word_i$ を含むページの検索ヒット数,

$hit_{key \cup word_i}$: キーワードか $word_i$ のどちらか含むページの検索ヒット数,

N : 抽出した名詞の数.

関連度 rel_i の算出には、新たに hit_{word_i} , $hit_{key \cap word_i}$ を取得する必要があるが、この取得には非常に時間がかかる。 hit_{word_i} , $hit_{key \cap word_i}$ を取得するためには、それぞれ抽出した名詞の数である N 回、合わせて $2N$ 回の web 検索が必要である。例え

ば、キーワードによる検索で上位 10 件の web ページを取得し、それぞれのページで 50 語の名詞が抽出されたとすると、全体で 500 語の名詞が抽出される。したがって、(2) 式により関連度を算出するには 1000 回の web 検索が必要となる。これを短縮するため、 $hit_{key \cap word_i}$ は式(3)により近似的に取得する。これは、 $hit_{key \cap word_i} / hit_{key}$ が、先に取得したキーワードを含む上位 M 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。この近似により、関連度の算出に要する検索回数を $1/2$ に減らすことができる。いくつかのキーワードについて近似を用いた方法と用いない方法で関連度を算出したところ、大きな違いは見られなかった。

$$hit_{key \cap word_i} \simeq hit_{key} \times page_{word_i | key} / page_{key}, \quad (3)$$

但し、

$page_{key}$: キーワードで取得した web ページ数,

$page_{word_i | key}$: 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 rel_i を算出する。この関連度 rel_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ーに提示する。

2.6 ユーザ選択をフィードバックとした地図データベースの更新

ユーザーは、地図データベース参照もしくは web 検索を用いた関連語抽出処理によって取得された索引候補語から、写真の内容に合った地名や施設名を選択する。選択された名称は、索引語として写真に対応付けられると共に、地図データベースへのフィードバックとして用いて、データベースの更新を行う。

図5は、ユーザー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。ユーザー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、索引語がユーザーによって選択された回数を示す頻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。この場合、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、以下のような処理を行う。まず、その索引語が既に地図データベースに登録されているか否かで処理を分岐する。索引語が地図データベースに登録されていない単語である場合、対応付けられ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う。索引語が既に地図データベースに登録されている単語である場合、地図データベースにおいて対応付けられた単語の位置情報と、ユーザーによる索引付加作業によって新たに対応付けられた写真のもつ位置情報を用いて、式(4)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、ユーザーによる

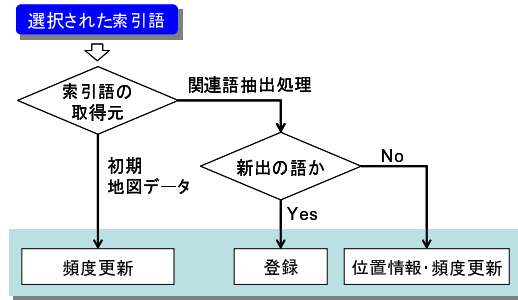


図5 ユーザ選択の地図データベースへのフィードバック

選択頻度についても更新を行う。

$$\begin{cases} lat_{new} &= (lat_{prev} \times freq_{prev} + lat_{sbj}) / freq_{new}, \\ lon_{new} &= (lon_{prev} \times freq_{prev} + lon_{sbj}) / freq_{new}, \\ freq_{new} &= freq_{prev} + 1, \end{cases} \quad (4)$$

但し,

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

lat_{prev}, lon_{prev} : DBに登録されている緯度・経度,

lat_{sbj}, lon_{sbj} : 推定被写体位置の緯度・経度,

$freq_{prev}, freq_{new}$: ユーザによる選択頻度.

索引語とこれを付加された写真の位置情報の対応関係を、地図データベースに保持しておくことで、次に同一の被写体が撮影された際の索引付加作業は効率化される。また、新たに得られる写真の位置情報を利用して、索引語に対応付けられた位置を更新する。つまり、同一被写体の写真が撮影され、ユーザによる索引付加作業のフィードバックが得られるごとに、地図データベース内の対応する索引語と位置情報からなるデータは、センサによる誤差の影響を低減するなどの目的から位置情報の平均化が行われる。また、よく使用される索引語の選択頻度は増加し、この2つの値に基づいた索引候補語の提示を行うことで、ユーザ入力の効率化が図られる。

3. 索引付加実験

3.1 プロトタイプシステム

提案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。プロトタイプシステムは、図6に示すようにクライアントとサーバからなる。クライアントは、写真と撮影時の位置・姿勢情報を取得するカメラ、GPS、ジャイロからなるセンサ付カメラと、これらの情報を記録し、サーバとの情報のやり取りを行うクライアントPCから構成される。システムに使用した機器類を表2に示す。サーバは、クライアントから送られる撮影位置・姿勢情報に基づき、格納されている地図データベースの参照およびweb検索を用いた関連語抽出処理を用いて、ユーザに写真の索引候補語を提示する。ユーザが図7に示すブラウザ上で動作する選択画面から、索引語を選択すると、写真と索引語は写真データベースに格納される。また、索引語をフィードバックとして用いた地図データベースの更新が行われる。

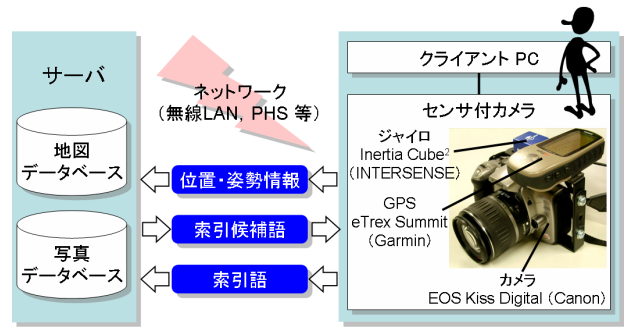


図6 プロトタイプシステムの構成

表2 プロトタイプシステムに用いた機器類

ジャイロセンサ	Inertia Cube ² (InterSense)	
GPS	eTrex Summit (Garmin)	
カメラ	EOS Kiss Digital (Canon)	
クライアントPC	LavieG (NEC)	
	web ブラウザ PHS	Mozilla FireFox 1.0 b-mobile (日本通信株式会社)
サーバPC	web サーバ	Apache 1.3.27
	データベース	PostgreSQL 7.3.2
	サーブレット	Tomcat 5.5.3, JDK5.0

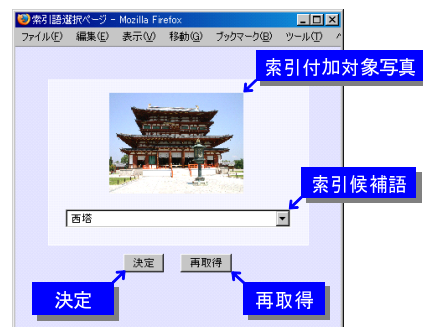


図7 索引語選択画面

プロトタイプシステムを用いた索引付加作業は、以下のように行われる。まず、ユーザがセンサ付カメラにより撮影を行うと、写真と撮影時の位置・姿勢情報がクライアントPCに記録される。記録した情報をクライアントPCから、ネットワークを介してサーバに送信し、索引付加作業を開始する。ユーザは、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し、サーバに送信する。サーバは、受信した情報から写真の被写体位置を推定し、サーバ内に設置された地図データベースを参照することで対応する位置の地名や施設名を索引候補語として取得する。これをクライアント側に返信すると、図7に示す索引語選択画面がユーザに提示される。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ボタンを押すことで、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と判断した場合には、ユーザが写真に付加したい索引語に最も関連すると思う単語をキーワードとして選択して再取得のボタンを押すことで、

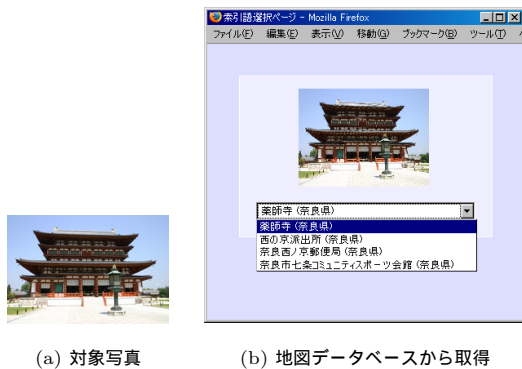


図 9 金堂 2



図 8 金堂 1

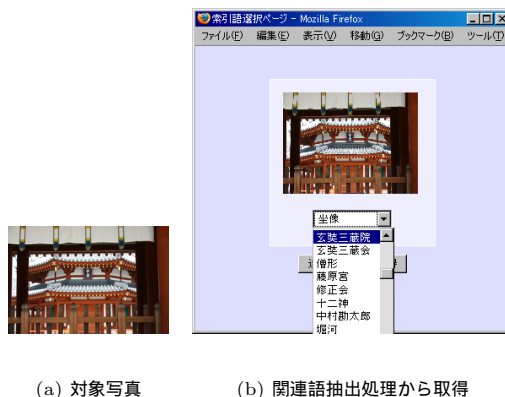


図 10 玄奘三蔵院

新たな索引候補語の取得がサーバに要求される．サーバは、クライアントから送られる単語から、web 検索を用いた関連語抽出処理を行う．この結果得られた関連語を再びクライアントに送信し、同様の選択作業がユーザによって行われる．ユーザによる選択作業が行われると、選択された索引語は写真に対応付けられると共に、サーバへと送信され、これをフィードバックとした地図データベースの更新処理が行われる．

初期地図データベースの構築には、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 [19] 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（図 4 において $M=10$ ）．また、形態素解析には日本語形態素解析システム「茶筌」[20] を使用した．

3.2 撮影した写真への索引付加実験

3.2.1 予備実験

予備実験では、システムが提案手法に基づいた動作をするか確認し、写真に対して適切な索引を付加できる場合とそうでない場合について検証した．図 8(a)~11(a) は、奈良の薬師寺において撮影したものである．これらの写真に対してプロトタイプシステムを用いた索引付加作業を行った．実験で用い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 8~11 の順に索引付加作業を行った様子について述べる．

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想定する．サーバは初期



図 11 興楽門

状態の地図データベースを保持しており、この写真とその撮影位置・姿勢情報を送信すると図 8(b) 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」をキーワードとして選択し、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 8(c) 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 3 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 4 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．

表 3 地図データベースの更新

input image	name	lat	lon	freq
図 8(a)	金堂	34 °40' 4.7"	135 °47' 3.5"	1
図 9(a)	金堂	34 °40' 6.0"	135 °47' 4.1"	2

表 4 抽出された「薬師寺」の関連語 (上位 20 語)

1	薬師尊像	11	東塔
2	薬師寺ホームページ	12	門扉
3	東院堂	13	西塔
4	老春手帳	14	白鳳時代
5	玄奘三蔵院	15	統天皇
6	フェノロサ	16	天武天皇
7	平山郁夫画伯	17	修二会
8	花会式	18	奈良駅
9	三重塔	19	金堂
10	光明皇后	20	催事名

太字は「薬師寺」内の施設名

次に図 9(a) 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(b) 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 9(a) のもつ位置情報を利用して更新される。結果、表 3 上段に示した「金堂」のレコードは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 43m である。使用した 2 枚の写真の推定被写体位置間の距離は約 85m 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などを検討する必要がある。

同様の処理によって、図 10(a) については図 10(b) 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 5 番目に提示されている。一方、図 11(a) については図 11(b) 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「興楽門」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「興楽門」という単語が含まれていなかったことによる。

予備実験における索引付加作業により、プロトタイプシステムが提案手法に基づいた動作をすることを確認した。具体的には、地図データベースから写真撮影時の位置・姿勢情報に基づく地名や施設名が取得されること、web 検索を用いた関連語抽出処理により、入力として与えた語に関連する語が取得されることを確認した。これらの処理により取得された単語を索引候補語としてユーザに提示することで、写真に対する半自動的な索引付加作業を実現することができた。また、ユーザが索引付加作業において選択した索引語をフィードバックとして、地図データベースの更新を行うことにより、同一の被写体に対する

作業が効率化されることを確認した。一方、提案手法では索引語を付加できない事例も確認されたため、これについては、索引候補語を取得する手法の改良や新たな手段の検討が必要であると考えられる。

3.2.2 実証実験

観光施設を対象として予備実験よりも多くの写真を撮影し、プロトタイプシステムを用いた索引付加作業によって写真に対してどの程度適切な索引を付加することができるかを確認するとともに、地図データベースの登録・更新について検証するための実験を行った。実験で索引付加の対象とした写真は、薬師寺において撮影した 9 施設、71 枚の写真であり、その一部を図 12 に示す。索引付加作業では、ユーザが写真の被写体であるそれぞれの施設に対して、その施設名を索引として付加することを想定する。地図データベースは初期状態とし、また、web 検索を用いた関連語抽出処理についての条件は、予備実験と同様である。

予備実験と同様の索引付加作業を、写真を撮影した順序で行った。表 5 は被写体とした各施設について、その名称と撮影枚数、web 検索を用いた関連語抽出処理で取得した際のユーザへの提示順位、地図データベースから取得した際の平均提示順位を示している。付加する索引語は索引候補語中、関連語抽出処理において 6 枚の写真を対象に平均 25.2 位、地図データベースからの取得において 55 枚を対象に平均 1.3 位に提示された。ほとんどの場合で、選択する索引語は候補語リスト中の上位にあり、ユーザによる選択作業は簡単なものであった。関連語抽出処理では、選択する索引語の一部はリストの下位にあり選択作業の負担は大きいものであった。しかし、索引語が一度地図データベースへ登録された後の作業ではリストの上位に提示されており、選択作業の負担は軽減された。実験を通して、地図データベースに一度登録された索引語については、ユーザにそれほど負担をかけることなく索引付加作業を行うことができた。但し、施設によって提示順位にはばらつきがあった。この原因としては、施設同士が近接している場合や施設が広がりをもつ場合に、対象施設の周囲の施設の方が上位に提示されることが挙げられる。写真撮影時にセンサから取得した位置・姿勢情報に含まれる誤差も原因の一つと考えられる。

9 施設中 3 施設 (71 枚中 10 枚) で適切な施設名を索引語として付加することができなかった。これは、web 検索による関連語抽出処理において取得した web ページ中に、それらの単語が含まれていなかったことが原因である。索引の付加できなかった施設は、観光施設としてはあまり注目されない施設であり、そのため、取得した web ページにも含まれていなかったと考えられる。同時に、一般の観光客が本システムを使用する際にも撮影される可能性は低いと考えられるため、実用上それほど問題にはならないと考えられる。

4. ま と め

本稿では、個人が撮影した写真を効率的に管理することを目的とし、従来から研究の行われている画像検索手法のうち、個人の写真について適しているとされるメタデータを利用して検



図 12 索引付加対象写真 (一部)

表 5 索引付加作業における索引語の提示順位

索引語	撮影枚数	関連語抽出処理における提示順位	地図 DB による平均提示順位
西塔	7	13	1.0
東塔	7	11	1.2
大講堂	11	100	1.5
玄奘三蔵院	18	5	1.8
金堂	12	19	1.4
東院堂	6	3	1.0
鐘楼	4	取得なし	-
與樂門	4	取得なし	-
南門	2	取得なし	-
平均	7.9	25.2	1.3

索を行う手法について検討した。また、メタデータのうち特に写真の撮影位置情報に着目し、これをユーザが閲覧や検索を行うのに適した被写体の地名や施設名等の索引として、写真に付加しておくための手法を検討した。

以上に基づき、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、写真に適切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。ここでユーザにより選択された索引語は、その位置に適切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。

提案システムのプロトタイプを用いて撮影した写真に対し位置・姿勢情報に基づく索引語の付加実験を行い、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できたと考えられる。

今後の課題としては、まず、提案手法で取得できていない地名や施設名の取得方法の検討が挙げられる。提案手法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って、索引語とする地名や施設名を取得しているため、取得できるのは、市販の地図データに含まれるか、web で取り上げられるような知名度のある対象となり、適用できる被写体は観光地等に限定されている。したがって、索引語取得方法の改良や手入力のための簡単なインタフェースの導入、また、写真の画像情報の利用等を検討する。次に、地図データベースに登録される候補語の位置情報について、実際の位置とずれが生じる点について検討する必要がある。また、撮影場所および利用者に関してより広範な評価実験を行うこと、構築した写真データベースにユーザがアクセスするためのインタフェースについて検討する。

文 献

- [1] A. W. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain: "Content-based image retrieval at the end of the early years", IEEE Trans. on Pattern Analysis and Machine Intelligence, **22**, 12, pp. 1349–1380 (2000).
- [2] R. C. Veltkamp and M. Tanase: "Content-based image retrieval systems: A survey", Technical Report TR UU-CS-2000-34 (revised version), Department of Computing Science, Utrecht University (2002).
- [3] "AltaVista Photofinder". <http://www.altavista.com/image/>.
- [4] "Google Image Search". <http://images.google.com/>.
- [5] "Adobe Photoshop Album". <http://www.adobe.com/>.
- [6] "Apple iPhoto". <http://www.apple.com/>.
- [7] R. Brunelli and O. Mich: "Efficient image retrieval by examples", Proc. 5th Working Conf. on Visual Database Systems, pp. 145–162 (2000).
- [8] M. Das, E. M. Riseman and B. A. Draper: "Focus: Searching for multi-colored objects in a diverse image database", Proc. 1997 Conf. on Computer Vision and Pattern Recognition (CVPR '97), p. 756 (1997).
- [9] E. D. Sciascio, G. Mingolla and M. Mongiello: "Content-based image retrieval over the web using query by sketch and relevance feedback", Proc. 3rd Int. Conf. on Visual Information and Information Systems, pp. 123–130 (1999).
- [10] J. R. Smith and S. fu Chang: "Querying by color regions using VisualSEEK content-based visual query system", Intelligent Multimedia Information Retrieval, MIT Press, pp. 23–41 (1997).
- [11] M. Beigi, A. B. Benitez and S.-F. Chang: "MetaSEEK: A content-based meta-search engine for images", Proc. SPIE 1998 Conf. on Storage and Retrieval for Image and Video Databases VI IST/SPIE 1998, Vol. 3312, pp. 28–30 (1998).
- [12] 相良, 砂山, 谷内田: "HTML テキストの重要文を用いた画像ラベリング手法", 電子情報通信学会論文誌 (D-I), **J87-D-I**, 2, pp. 145–153 (2004).
- [13] Y. A. Aslandogan and C. T. Yu: "Multiple evidence combination in image retrieval: Diogenes searches for people on the web", Proc. 23rd Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 88–95 (2000).
- [14] B. Shneiderman and H. Kang: "Direct Annotation: A drag-and-drop strategy for labeling photos", Proc. Int. Conf. on Information Visualization, pp. 88–95 (2000).
- [15] M. Naaman, Y. J. Song, A. Paepcke and H. Garcia-Molina: "Automatic organization for digital photographs with geographic coordinates", Proc. 2004 Joint ACM/IEEE Conf. on Digital Libraries, pp. 53–62 (2004).
- [16] K. Toyama, R. Logan, A. Roseway and P. Anandan: "Geographic location tags on digital images", Proc. 11th ACM Int. Conf. on Multimedia, pp. 156–166 (2003).
- [17] J. Electronics and I. T. I. Association (JEITA): "Exchangeable image file format for digital still cameras: Exif version 2.2" (2002).
- [18] 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, **153**, 8, pp. 57–64 (2003).
- [19] "Google Web API". <http://api.google.com/>.
- [20] 松本: "形態素解析システム「茶筌」", 情報処理, **41**, 11, pp. 1208–1214 (2000).