

L_057

地理情報データベースによる半自動写真キャプションニングの効率向上 Improvement of Semi-automatic Location-based Photo Captioning Using Geographical Database

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

1. はじめに

近年、ユーザの現在位置に基づく情報提供サービスが盛んに行われ、これを提供するカーナビや携帯電話をはじめとしたモバイル機器も広く普及してきている。この位置情報に基づく情報提供サービスにおいて、情報コンテンツの作成は人手で行われることが多く、提供する情報の充実には、その作成にかかるコストの削減が必要といえる。また、情報コンテンツは従来、それに携わる業者により作成・提供されてきたが、個人がwebページやブログなどにより情報を発信する現在では、個人が地域の店舗情報等を発信しているサイトも多く見受けられる。その際、住所や緯度・経度、これを視覚的に提示する地図画像に加え、店舗の外観や概要を示す写真のある情報コンテンツは、情報を受け取るユーザにとって、その対象を理解する上で有用である。

Naamanらは、撮影位置情報(緯度・経度)とその場所の名称(地名、施設名など)の付加された写真を格納したデータベースを予め用意し、新たに撮影した写真について、その撮影位置情報に基づいてデータベースを参照し、類似の位置で撮影された写真の場所の名称を取得する手法[1]を提案している。藤田らはさらに、撮影方向を考慮して写真に対応する名称を取得する手法[2]を提案している。これらの手法は位置に関する名称を写真に対して与えるものであるが、いずれも、予め写真とその位置情報および位置に関する名称の組を保持しておく必要があり、また、全自動で名称を決定するため、名称が誤っていたり、ユーザの意図とは異なる名称が写真に対して付加されるおそれがある。

一方、我々は従来から地理情報データベースとweb検索を用いた半自動写真キャプションニングシステムを提案している[3]。このシステムは位置情報に基づく名称(キャプション)の候補をユーザへ提示し、ユーザがこの中から写真に合ったキャプションを選択するという半自動によるキャプションニングを行うものである。さらにこの作業を写真を撮ったその場で行うことで、ユーザの手間を軽減して正しいキャプションを付加する。しかし、地理情報データベースに含まれるデータの空間的な位置が点で表されているために、写真撮影時の位置に基づいてデータベースを参照し、キャプション候補を取得する際、対象の大きさや対象との位置関係によっては、ユーザに正しいキャプション候補を示すことができない、正しいキャプションを候補中の上位に示すことができないといった問題があった。また、web検索を用いた関連語抽出処理において取得されるキャプション候補に、「場所」に関連のない語が含まれるという問題が残されていた。そこで本稿では、キャプションニング効率を向上させるための改良について述べる。改良点は、(1)地理情報データベースに含まれるデータである施設や建物の位置を緯度・経度で表される点ではなく、そのオブジェクト

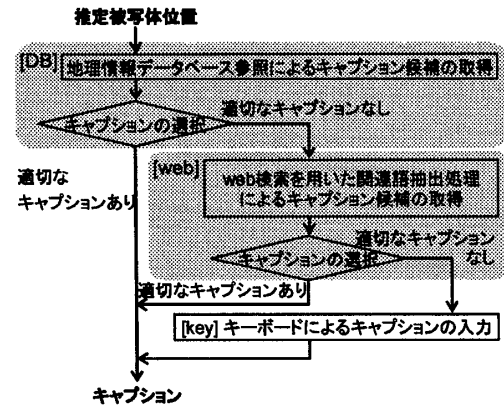


図1: キャプション入力の流れ

の占める領域として与えること、そして、(2) 関連語抽出処理において単語を語彙により分類した辞書を参照することである。以上によりキャプション時にユーザに対して示されるキャプション候補の精度を向上させることで、ユーザによるキャプションニング作業効率の向上を図る。さらに撮影した写真に対してこれらを適用した結果を示し、考察する。

2. 地理情報データベースとweb検索を用いた撮影位置・姿勢情報に基づく写真キャプションニング

2.1 キャプションニングの概要

我々は、写真に対して撮影時の場所に基づくキャプションを付加する枠組みを提案している[3]。ユーザによるキャプションニングは、以下の流れで行われる。

- (1) ユーザはGPSやジャイロ、コンパス等のセンサとカメラを用い、撮影位置・姿勢情報付きの写真を取得する。
- (2) キャプション候補生成のため、撮影時の位置・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。
- (3) ユーザは対話的にキャプションを入力する。入力における処理の流れを図1に示す。各処理の概要は以下の通りである。

(3-1) [DB] 地理情報データベースを用いた候補提示による入力

(3-1-1) 推定した被写体位置を用いて地理情報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。

(3-1-2) 取得した地名や施設名は写真のキャプション候補として提示され、ユーザは候補の中から被写体に適切なキャプションを選択する。

ステップ3-1-2で提示された候補に適切なキャプションが含まれていないと判断した場合、以下の処理を行う。

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

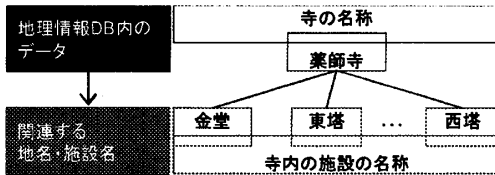


図 2: 関連する地名・施設名の取得

(3-2) [web] web 検索による関連語抽出処理を用いた候補提示による入力

(3-2-1) ユーザはステップ 3-1-2 で提示されたキャプション候補から付加したいと考えるキャプションに最も関連すると思われる語(以下、キーワード)を選択する。

(3-2-2) キーワードを用いて web 検索を行い、その結果から関連語を抽出する。例えば、図 2 に示すように、「薬師寺」という施設名をキーワードとして選択した場合、「薬師寺」内の建物である「金堂」、「東塔」といった施設名が抽出されるといった、より詳細なレベルの名称等の関連する名称を取得する。

(3-2-3) 抽出した関連語は新たなキャプション候補として提示され、ユーザは候補から被写体に適切なキャプションを選択する。

ステップ 3-2-3 で提示された候補に適切なキャプションが含まれていないと判断した場合、以下の手入力を行う。

(3-3) [key] キーボードによる入力

ユーザはキーボードにより被写体に適切なキャプションを入力する。

(4) ユーザにより入力されたキャプションを写真のメタデータとして画像ファイル内に書き込む。

(5) 入力されたキャプションを地理情報データベースへのフィードバックとして更新を行う。これにより、ユーザに提示されるキャプション候補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。

上記の手順でキャプション付きの画像ファイルが生成され、ユーザはキャプションに基づく閲覧や検索を行うことが可能となる。手法の詳細については参考文献 [3] を参照されたい。

2.2 キャプションニングの効率向上

2.1 節で示した枠組みでは、ユーザが写真を撮るとキャプション候補が提示され、ユーザはその中から写真に合ったキャプションを選択する。この選択作業の効率を向上させるため、地理情報データの空間領域への割り当て、語彙に基づく関連語のフィルタリングの二点の変更を加えた。これらの処理について以下で説明する。

2.2.1 地理情報データの空間領域への割り当て

従来の枠組みにおいて地理情報データは一組の緯度・経度による点として与え、この点に施設や建物の名称を対応付けたものとしていた。しかし実際には、施設や建物は空間的な広がりをもつため、撮影した写真の推定被写体位置と点として与えられた地理情報データの位置が一致することは稀で、近傍の地理情報データから尤もらしいデータを推定することになる。しかしこの方法では、複数の対象が近接している場合や対象の面積が広い場合、

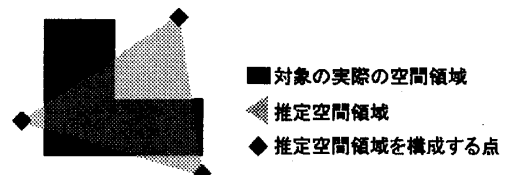
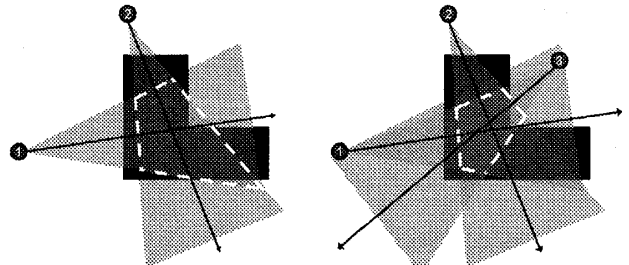


図 3: 地理情報データベースへの空間領域の登録



(a) 2枚の写真からの推定

(b) 3枚の写真からの推定

● N番目の撮影位置 — 撮影方角 ◀ 視野 ■ 推定空間領域

図 4: 地理情報データに対応する空間領域の推定

正しい地理情報データを推定することは困難になる。そこで、地理情報データを空間領域に対応付けるものとし、これに伴い必要となる処理やデータベースの構成について述べる。まず、空間領域と視野について説明する。

空間領域: 地理情報データとして登録される場所・施設・建物等が現実空間で占める二次元的な領域を指す。

視野: 写真に記録されたと推測される領域であり、撮影位置・姿勢情報と焦点距離などのカメラパラメータから求める。方位角(奥行)方向の範囲は不明であるため、予め一定値を設定する。

次に、処理とデータベースの構成について述べる。

地理情報データベースへの空間領域の登録

データベース内の地理情報データはその空間領域を多角形として表すものとし、構成する複数点をデータとして登録する。具体的には、図 3 に示すように対象の推定空間領域を構成する点の緯度・経度の組を格納することとなる。

地理情報データに対応する空間領域の推定

新たな写真が取得され、地理情報データに対応付けられた空間領域を更新する際、新たな写真を撮影したときの位置・姿勢情報やカメラパラメータを用いて空間領域を推定する。図 4 に空間領域の推定例を示す。図 4(a) は、ある対象の写真が 2 枚得られている場合の空間領域の推定結果を表す。2 枚の写真それぞれを撮影した際の撮影位置・姿勢、焦点距離などのカメラパラメータから視野を推定し、2 枚の写真の視野の積を対象の空間領域とし、これを構成する多角形の点を地理情報データとしてデータベースに格納する。また、図 4(b) は 3 枚の写真について同様の推定を行った際に得られる対象の空間領域を示している。

空間領域に基づくキャプション候補の取得

空間領域として定義された地理情報データを用いて、写真に対するキャプション候補を取得する。従来法 [3] では、写真のメタデータ Exif に含まれる被

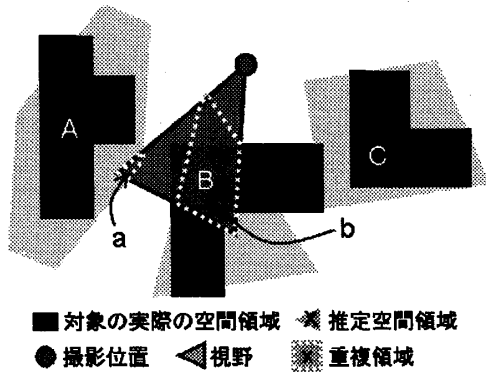


図5: 空間領域に基づくキャプション候補の取得

写体距離に基づき、被写体位置の推定を行っていた。しかし、被写体距離は被写体までのおよその距離を示すもので、正確な値が得られない場合があるため、空間領域に基づくキャプション候補の取得には被写体距離を用いず、以下で述べる処理を行う。

- (1) 対象写真の撮影位置からしきい値以内の距離に位置する地理情報データを取得する。
- (2) 対象写真の撮影位置・姿勢情報、カメラパラメータから、撮影時の視野を推定する。
- (3) 地理情報データの空間領域と推定した視野との重複領域の大きさを求める。
- (4) 求めた重複領域の大きな順に地理情報データを並べ替える。重複のないデータについては距離の近い順に並べ替える。この結果をキャプション候補としてユーザへ提示する。

キャプション候補の取得例を図5に示す。地理情報データの推定空間領域A, B, CのうちA, Bについて、それぞれ視野との重複領域a, bが存在する。重複領域bがaよりも大きいことから、ユーザに提示されるキャプション候補はB, A, Cの順となる。

2.2.2 単語の意味属性に基づく関連語のフィルタリング

web 検索による関連語抽出処理において、日本語語彙大系 [4] を参照することで「場所」との関連が低いと推測される語を除く。

日本語語彙大系は、日本語語彙の意味的用法を約3,000の意味属性を用いて定義しており、一般名詞のほかに使用頻度の高い固有名詞や新聞記事に現れる時事用語などを含む30万語を収録している。意味属性は階層となっており、例えば、固有名詞の意味属性「地名」の下位に「施設名」があるといった構造となっている。

そこで、従来提案している関連語抽出処理により得られた語の意味属性を日本語語彙大系を参照することで取得し、取得した意味属性に基づき、ユーザによるキャプションにおいて必要となる「場所」に関連する語のみを抽出する。具体的には、語の意味属性を日本語語彙大系から取得し、1つ以上得られた意味属性の中に表1に挙げた「場所」に関連する意味属性を含む語をキャプション候補として抽出するものとする。また、語が複数の単語から構成されている場合、日本語語彙大系を参照しても意味属性が得られない場合がある。この場合には、対象の語を形態素解析した結果得られる単語で意味属性を取得するものとする。以下にその例を示す。

表1: フィルタリングにより取得する語の意味属性 (日本語語彙大系の分類による)

意味属性体系	意味属性
一般名詞	場所, 組織
固有名詞	組織名, 地名

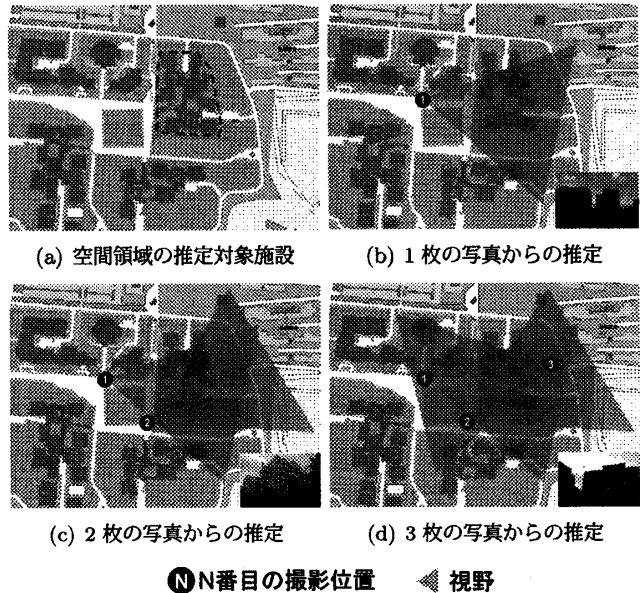


図6: 空間領域の推定例

キャプション候補として使用する: 「東塔」

「東」の意味属性として「町」「その他の区」「大字(その他)」「方角」「村」「名(男)」「姓」を持ち、「塔」の意味属性として「家屋(本体)」「大字(その他)」「施設」を持つため、「場所」に関連する語と判定し、キャプション候補として使用する。

キャプション候補として使用しない: 「略称」

「略称」の意味属性は「呼称」「名称」であるため、「場所」に関連しない語と判定し、候補から除去する。

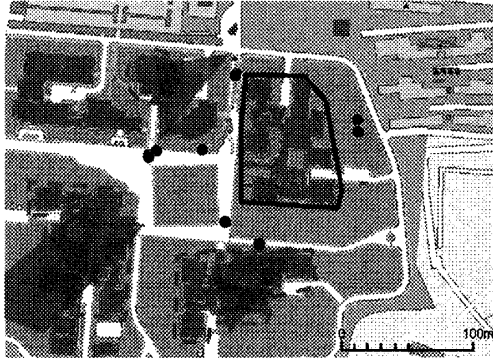
3. 実験

2.節で述べた方法に基づいてシステムを構築し、キャプションの効率について評価した。なお、初期状態のデータベースには、アルプス社「プロアトラス W2」施設データを使用し、これは、各施設データについて緯度・経度による点を位置情報として与えるものである。

3.1 地理情報データの空間領域への割り当て

3.1.1 地理情報データに対応する空間領域の登録・更新

ある施設を撮影した複数の写真を用いて空間領域を推定した例を図6に示す。ここで対象とするのは図6(a)に示す施設とする。1枚目の写真の撮影位置・姿勢及びカメラパラメータの情報から、図6(a)の視野が推定される。次に、2枚目の写真の情報から、図6(b)の撮影位置2の視野が推定され、2枚の写真の視野の積となる領域が、対象施設の空間領域として地理情報データベースに設定される。同様に、3枚目の写真の情報から図6(c)の推定が行われ、対象施設の空間領域が更新される。



● キャプション対象写真の撮影位置 - - 近接施設の実際の空間領域
 — キャプション対象施設 ■ 近接施設の推定空間領域

図 7: 地理情報データの空間領域と対象写真の撮影位置

表 2: 空間領域に基づくキャプション候補取得による提示順位の変化

	従来法	提案法
平均提示順位	1.6	1.0

3.1.2 空間領域に基づくキャプション候補の提示

地理情報データを1組の緯度・経度からなる点で表す従来法 [3] と、空間領域で表す提案法によってユーザへ提示されるキャプション候補を比較した。実験において対象としたのは、指定した1つの施設を、その周辺の異なる地点・方向から撮影した10枚の写真である。地理情報データベースは、前述したように市販の地図ソフトウェアに収録されたデータが登録され、さらに提案法については、キャプション対象とは別に取得した写真により、対象施設と近接する3施設の空間領域をそれぞれ写真2枚から推定し、これを登録した状態とした。図7はこれらの位置関係を示したものである。写真のキャプションとして適切な候補(対象の施設名)が、システムにより提示されたキャプション候補中の何位であったかを、キャプション効率の指標として表2に示す。結果から、提案法は適切なキャプションを従来よりも上位に提示していることが確認された。

3.2 単語の意味属性に基づく関連語のフィルタリング

地理情報データベース内に初期状態で含まれるデータのいくつかをキーワードとして、2.2.2節で述べた方法により関連語の抽出を行った。

表3に「薬師寺」をキーワードとした場合に、単語の意味属性に基づき、「場所」に関連する語のみを取得するフィルタリングを行った場合と行っていない場合について、その結果を示す。結果からフィルタリングにより「場所」に関連しない語が除かれていることがわかる。例えば、フィルタリングなしの場合には、「拝観」がキャプション候補として取得されていたが、意味属性に基づくフィルタリングを行った結果には含まれていない。フィルタリングありの結果に含まれている「持統天皇」という語については一般には「場所」に関連しないと考えられるものの、「天皇」が「大字(その他)」という「場所」に関連する意味属性を持つためにキャプション候補として取得されている。また、フィルタリングなしの結果に含まれる「薬師尊像」という語については、被写体を表すキャプションとしたい場合も想定されるため、フィルタリングの対象とする意味属性をユーザに選択してもらいなどの検討が必要であると考えられる。

表 3: 意味属性に基づく関連語のフィルタリング結果 (キーワード「薬師寺」, 上位10語)

順位	フィルタリングなし	フィルタリングあり
1	大宝蔵殿	大宝蔵殿
2	東塔	東塔
3	西塔	西塔
4	三重塔	三重塔
5	拝観	東大寺
6	東大寺	東院堂
7	再建	花会式
8	薬師尊像	西の京
9	東院堂	持統天皇
10	裳階	大講堂

表 4: フィルタリングによるキャプション候補数の変化

キーワード	フィルタリング	
	なし	あり
薬師寺	517	167
奈良先端科学技術大学院大学	201	177
奈良県庁	1012	374
平均	577	239

表4は意味属性に基づくフィルタリングによるキャプション候補数の変化を、いくつかのキーワードについて示したものである。この処理により候補数は多くの場合半分以上に削減され、「場所」に基づくキャプションの効率は向上したと考えられる。

4. まとめ

本稿では、我々が従来から提案している撮影位置・姿勢情報に基づく半自動写真キャプションの枠組みにおいて、ユーザによるキャプションの効率向上のために、以下の二点の改良を行った。

- 地理情報データベースに含まれるデータである施設や建物の位置を緯度・経度で表される点ではなく、そのオブジェクトの占める空間領域として与える。
- 関連語抽出処理において単語を語彙により分類した辞書を参照する。

以上によりキャプション時にユーザに対して示されるキャプション候補をより写真に即したものとすることができた。今後の課題としては、地理情報データに対応する空間領域は、取得される写真が増加するほど複雑になるという問題があり、これに対処する領域推定方法、データベースへの格納方法を検討する必要がある。

参考文献

[1] M. Naaman, A. Paepcke and H. Garcia-Molina: "From where to what: Metadata sharing for digital photographs with geographic coordinates", Proc. 10th Int. Conf. on Cooperative Information Systems, pp. 196-217 (2003).
 [2] 藤田, 有川: "空間関係に基づく写真と注釈のデータ相補発展", 情報処理学会論文誌, 47, 1, pp. 63-76 (2006).
 [3] K. Iwasaki, K. Yamazawa and N. Yokoya: "An indexing system for photos based on shooting position and orientation with geographic database", Proc. IEEE Int. Conf. on Multimedia and Expo, CD-ROM, No. 1259 (2005).
 [4] NTTコミュニケーション科学基礎研究所: "日本語語彙大系 CD-ROM 版", 岩波書店 (1999).