

撮影位置・姿勢情報とネットワーク共有型データベースを用いた 写真キャプションニング

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、写真を撮影する機会は日常化してきている。しかし写真を簡単に管理する方法は少なく、データが未整理のまま利用されない、利用しようとしても必要な写真をなかなか見つけられないということも多い。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であるキャプション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。本稿では、ユーザ間で共有する地理情報データベースとweb検索を用いた関連語抽出処理により、撮影位置・姿勢情報に基づくキャプションを撮影したその場で半自動的に生成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評価実験について考察する。キーワード 撮影位置・姿勢情報, 写真キャプションニング, 地理情報データベース, web検索, 関連語抽出

Captioning Photos Using Networked Shared Database 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 easily, and unorganized photos are not used or are difficult to avail. 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 captions manually. It is also difficult to add captions intended by a user automatically. We have proposed a semi-automatic photo captioning system that enables users to generate captions easily. The proposed system generates captions using geographical database and web retrieval based on shooting position and orientation information. In this report, we evaluate a prototype system by some experiments.

Key words shooting position and orientation, photo captioning, 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、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。近年、写真に対するメタデータの付加が一般的になりつつある。例えば市販のデジタルカメラではメタデータの標準規格として Exif [1] があり、撮影日時やカメラパ

ラメータ、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。これに伴い、メタデータを利用して写真の検索を行う手法 [2] やメタデータとして写真に付加するキャプションを簡単に生成する手法 [3, 4] が提案されている。

一般に、撮影した写真を閲覧・検索する場合、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通常、写真のメタデータとして付加されて

おり、日時に基づく閲覧・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限られた人物のみが撮影されていると考えられる。そこで、顔認識手法を用いたシステム [5] や写真の人物にラベル付けを行うインタフェース [4,6] が提案され、人物に基づく閲覧・検索を実現している。出来事についてはスケジュール管理を行うソフトウェアを参照し、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 [7] 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 [8] や GPS を利用して位置情報を取得するシステム [7,8] 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、閲覧や検索においてユーザが必要とする情報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、場所に基づく閲覧や検索を行うには課題がある。

我々は「場所」に着目し、地理情報データベースと web 検索を用いた関連語抽出処理により撮影位置・姿勢情報に基づくキャプションを半自動的に生成するシステムを提案している [9]。キャプションの候補は、予め用意された地理情報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キャプション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を取得しユーザに提示する。ユーザに選択されたキャプションは、その位置に適切な語であると見なし、これを地理情報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が変化し、ユーザの選択作業は効率化される。本稿では、プロトタイプシステムを用いてユーザによる評価実験を行った結果について考察する。

以降、2 章では撮影位置・姿勢情報とネットワーク共有型データベースを用いた写真キャプションの概要を、3 章では提案するシステムのプロトタイプを用いた評価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・姿勢情報とネットワーク共有型データベースを用いたキャプションの概要

図 1 に我々が提案した撮影位置・姿勢情報をもつ写真に対するキャプション [9] の処理の流れを示す。各処理の詳細は以下の通りである。

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。
3. 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。
4. 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ユーザに提示され、ユーザは提示された候補の中か

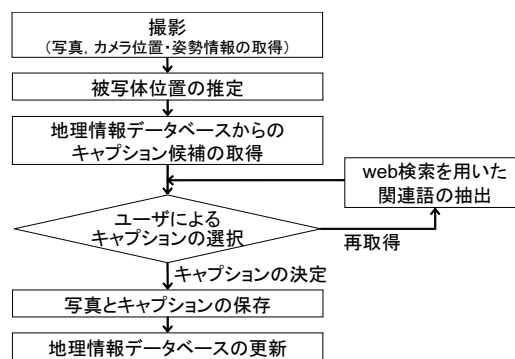


図 1 撮影位置情報付き写真へのキャプションング

ら被写体に適切なキャプションを選択する。

処理 4 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、

5. ユーザは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。
6. 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、処理 5 において地理情報データベースによる候補から「薬師寺」という施設名のデータを選択した場合「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。
7. 抽出した関連語は新たなキャプション候補としてユーザに提示され、ユーザは提示された候補の中から被写体に適切なキャプションを選択する。

処理 7 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、

8. 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。

処理 4, 7, 8 のいずれかによりキャプションが決定された場合、

9. ユーザにより選択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。
10. 選択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 [9] を参照されたい。

3. キャプションシステムの評価実験

3.1 プロトタイプシステム

プロトタイプシステムは図 2 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、電子コンパス付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通信を行う PC から構成される。サーバはユーザ間で共有する地理情報データベースを保持し、クライアントから送られる撮影位置・姿勢情報とカメラパラメータに基づき、格納されている地理情

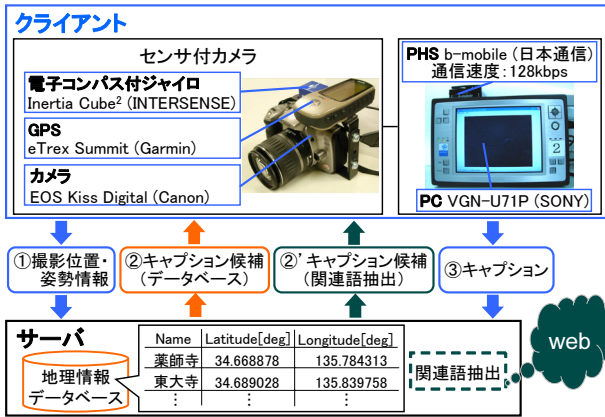


図 2 プロトタイプシステムの構成

表 1 サーバに用いたソフトウェア及びデータ

HTTP 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
地理情報データ	アルプス社「プロアトラス W2」施設データ
web 検索エンジン	Google API [10]
形態素解析	日本語形態素解析システム「茶筌」[11]

報データベースの参照及び web 検索を用いた関連語抽出処理を用いてキャプション候補を生成する。表 1 にサーバに用いたソフトウェア及びデータを示す。プロトタイプシステムを用いたキャプションングでユーザが行う作業は以下の通りである。

1. センサ付カメラで対象を撮影する。
2. PC に表示されるインタフェースから以下の方法でキャプションを入力する (図 3(a) 参照)。

(DB) 地理情報データベースを用いた候補提示による入力
地理情報データベースの参照によりサーバから提示された候補からキャプションを選択する (図 3(b) 参照)。

(DB) 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、

(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
再取得のチェックボックスにチェックを入れ、ユーザが写真に付加したいキャプションに最も関連すると思う語をキーワードとして選択する。サーバから提示された web 検索を用いた関連語抽出処理により得られた候補からキャプションを選択する。

(web) 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、

(key) キーボードによる入力
ユーザがキーボードにより適切なキャプションを入力する (図 3(c) 参照)。

キャプションの入力が終わると、キャプションとセンサから取得された撮影時の位置・姿勢情報は写真のメタデータ部分に書き込まれ、キャプション付きの画像ファイルが作成される。

3.2 評価実験の概要

提案手法におけるキャプション入力方法、(DB) 地理情報データベースを用いた候補提示による入力、(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、(key) キーボー

ドによる入力について評価する実験を行った。実験では 3.1 節で述べたプロトタイプシステムに加えて、図 4(a) に示す、より簡便な機器構成のシステムについても同様の評価を行った。このシステムはデジタルスチルカメラ (以下、DSC) に換えて USB カメラ (Logitech QV-700N) を用いたものであり、被写体距離を含む Exif 情報はない。したがって提案手法における被写体位置の推定ができないため、このシステムでは撮影位置に基づいて候補を提示した。また、DSC を用いるシステムがカメラのシャッターボタンを押して撮影を行うのに対し、USB カメラを用いるシステムは、図 4(b) のように PC に表示されるプレビュー画面を直接クリックすることで撮影を行った。

表 2 は評価実験に用いた 6 つのシステムである。システムは使用するカメラとキャプション入力方法の組み合わせがそれぞれ異なり、機器構成及びキャプション入力方法の違いによる比較を行った。評価項目は以下の通りである。

- ユーザが使用したキャプション入力方法 (各入力方法でキャプションを付加された写真の枚数)
- ユーザによるキャプションの入力にかかる時間
- ユーザにキャプション候補を提示する入力方法について、候補の中で適切なキャプションが提示された順位
- ユーザのシステム使用により地理情報データベースに登録・更新されるデータの位置 (緯度・経度)
- ユーザによる使いやすさに関する主観評価

6 つのシステムにおいて使用するカメラとキャプション入力方法、地理情報データベースは以下の通りである。

- カメラ
 - (1) システム 1-1, 1-2, 1-3 : DSC
 - (2) システム 2-1, 2-2, 2-3 : USB カメラ
- キャプション入力方法
 - 3.1 節で述べた提案システムどおりの方法、もしくはその一部を用いる方法とした。
 - (1) システム 1-1, 2-1:キャプション入力方法 (DB), (web), (key) をキャプションが入力できるまで順に行う。
 - (2) システム 1-2, 2-2:キャプション入力方法 (DB), (key) をキャプションが入力できるまで順に行う
 - (3) システム 1-3, 2-3 : キャプション入力方法 (key) のみを行う。入力の際、最近入力した履歴を候補として提示し、キー入力に従って候補を絞り込む機能を付加する。
- 地理情報データベース

データベースは各システムごとに別箇に用意する。ユーザがシステムを使用することでデータが追加・更新され、システムを最初に使用したユーザ以外はこの更新された状態のデータベースを使用する。

- (1) システム 1-1, 1-2, 2-1, 2-2 : 初期状態は市販の地図ソフトウェア (アルプス社「プロアトラス W2」) に収録された施設データを登録した状態とする。
- (2) システム 1-3, 2-3 : 使用しない。

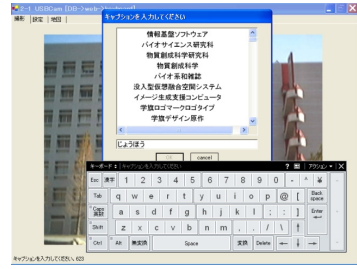
これらのシステムを用いて、20 代から 30 代の被験者 12 名が各システムにつき 4 枚、計 24 枚、実験全体では 288 枚の写真撮影した。写真は大学構内の建物を撮影したもので、4 枚



(a) ユーザによる入力の様子



(b) 候補選択



(c) キーボード入力

図 3 キャプションの入力



(a) 機器構成



(b) 撮影の様子

図 4 USB カメラを用いたプロトタイプシステム

表 2 評価実験に用いたシステムの構成

システム No.	カメラ	キャプション入力方法
1-1	DSC	DB web key
1-2		DB key
1-3		key
2-1	USB カメラ	DB web key
2-2		DB key
2-3		key

DB: 地理情報データベースを用いた候補提示による入力
 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
 key: キーボードによる入力

の内訳は以下の通りである。

- 建物 A (情報科学研究科): 2 枚
- 建物 B (バイオサイエンス研究科): 1 枚
- 建物 C (物質創成科学研究科), D (ミレニアムホール), E (ゲストハウスせんたん), F (サイエンスプラザ): うち任意の 1 枚

図 5 に、これらの建物の位置と大学周辺において初期状態の地理情報データベースに登録されているデータの位置を示す。建物 D のサイエンスプラザの名称は登録されているが、他の建物については名称が登録されていない。なお、ユーザが対象の建物を撮影する位置は対象から近くても遠くても構わないとし、また各建物にキャプションとして付加する名称は予め指定した。システムの評価順序はユーザごとにランダムに変更した。

3.3 結果と考察

被験者が撮影した写真の一部を図 6 に示す。対象を遠景・近景、また、さまざまな角度から撮影していることがわかる。

表 3 は、各評価システムにおいて、(DB) 地理情報データベースを用いた候補提示による入力、(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、(key) キーボードによる入力のそれぞれの入力方法でキャプションを付加した写真の枚数、その際の平均入力時間、また、候補を提示して選択する



□ : 初期状態で地理情報データベースに登録されているデータの位置
 ○: 撮影対象の位置

図 5 初期状態の地理情報データベースのデータ及び撮影対象の位置

入力を行う方法については、選択されたキャプションの候補中での順位の平均を示したものである。なお、入力時間はユーザが撮影を開始してからキャプションを入力するまでの時間であり、(web) 及び (key) による入力についてはその前に行う (DB) や (web) による入力で失敗した際に要した時間を含む。

入力方法 (DB) によってキャプションの入力を行った場合 (システム 1-1, 1-2, 2-1, 2-2 の DB), いずれのシステムでも 10 秒程度で入力ができている。DSC で撮影を行うシステムが USB カメラのシステムよりも入力に時間がかかっているのは、DSC から PC への画像の転送等によって 3 秒程度、ユーザへのキャプション提示に時間がかかるためであり、これを考慮すると入力時間は 4 つのシステムで同程度と考えられる。また、提示順位についても平均で 3 位までに提示されており、ユーザの選択作業は簡単なものであったと考えられる。DSC を用いたシステムでは Exif の被写体距離を用いて被写体位置の推定を行って、これを地理情報データベースからの候補提示やデータベースのキャプション登録位置の更新に使い、一方、USB カメラを用いたシステムでは Exif の被写体距離が得られないため撮影位置を用いて処理を行った。今回の実験では、DSC と USB カメラを用いたシステム間であまり差が見られなかった。これは被写体位置の推定に用いている Exif の被写体距離が、ピントの合った位置までの距離をおおよそ示すもので、撮影の際に対象にピントが合っていない場合や被写界深度が深い場合などに正しい値を示さないことなどが原因と考えられる。図 7(a), 7(b) にシ



図 6 撮影された写真（一部）

表 3 キャプション入力作業の評価

システム No.	1-1			1-2		1-3	2-1			2-2		2-3
	DB	web	key	DB	key	key	DB	web	key	DB	key	key
キャプションを付加した枚数	43	3	2	44	4	48	44	3	1	45	3	48
平均入力時間 [秒]	10	44	113	10	52	26	8	46	108	9	38	22
平均提示順位	1.8	66.3	-	2.9	-	-	2.4	66.3	-	2.8	-	-

注：各システムの総枚数は 48 枚



(a) システム 1-1：撮影位置と DB 登録位置 (b) システム 1-1：推定被写体位置と DB 登録位置 (c) システム 2-1：撮影位置と DB 登録位置

x：データベース登録位置 o：撮影位置または推定被写体位置 A：撮影対象の実際の位置

図 7 撮影による地理情報データベースへのデータの登録：建物 A（情報科学研究科）

システム 1-1 を使用して建物 A（情報科学研究科）を撮影した際の撮影位置と推定被写体位置，そして実験後の地理情報データベースにおけるキャプションの登録位置を示す。Exif の被写体距離は実際の被写体までの距離よりも小さいことが多く，結果として，推定被写体位置が実際よりも撮影位置に近くなっていることがわかる。被写体位置の推定を行うシステム 1-1 は，図 7(c) に示す推定を行わないシステム 2-1 よりも実際の位置に近く登録されているものの，前述したように撮影位置に近い値となっている。

入力方法 (DB) による入力に失敗し (web) による入力を行った場合 (システム 1-1, 2-1 の web)，入力には 45 秒程度かかっている。この方法により 3 枚の写真にキャプションが付加されており，その際の提示順位はそれぞれ 4 位 (バイオサイエンス研究科)，5 位 (物質創成科学研究科)，190 位 (情報科学研究科) であった。このため，前の 2 枚については簡単に選択できていたものの，後の 1 枚は選択に時間がかかっており候補として挙がっていても簡単に選択できていない。この点については携帯電話などでのテキスト入力に使用されている予測変換との組み合わせにより，効率的な入力に改良できると考える。

2 つの候補提示方法で写真に合ったキャプションを提示でき

ない場合には，はじめからキーボードで入力する方法 (システム 1-3, 2-3) が最も速い入力となった。

図 8 は，システム 1-1 を使用した際のユーザの作業時間である。建物 F 以外についてはデータベースにないキャプションを付加したため，1 枚目の撮影では作業に時間がかかっている。しかし，キャプションがデータベースに登録された状態である，建物 F の撮影とその他の建物の 2 枚目以降の撮影は 10 秒程度でキャプションが行われていることがわかる。また，実験終了時におけるシステム 1-1 で使用したデータベースへの登録・更新結果を図 9 に示す。初期状態で登録されていたデータに加えて，被験者のシステム使用により新たなデータが登録されている。データの登録位置は前述したように，被写体位置よりも撮影位置に近い位置となっており，データに対応する位置の登録方法については検討が必要である。具体的には，撮影位置と方位を撮影ベクトルとして，同じ被写体を撮影した複数枚の写真における撮影ベクトルの延長上で交わる点を被写体位置とする方法が考えられる。

次に，ユーザによるシステムのアンケート評価結果を表 4 に示す。評価は，撮影・キャプションングを通して使いやすいと感じた順に順位を付けるものとし，複数のシステムで順位の重

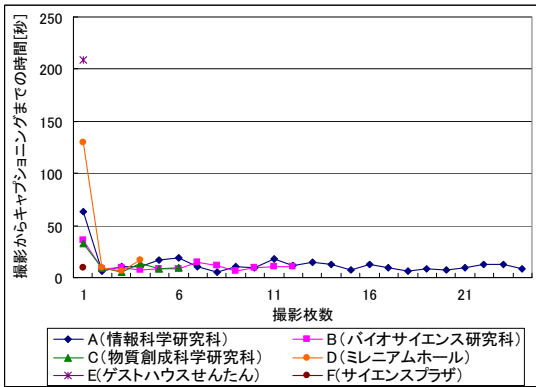


図 8 システム 1-1 使用時におけるユーザの作業時間



図 9 システム 1-1 のデータベースの登録・更新結果

表 4 ユーザによるシステムの使いやすさに関する主観評価

システム No.	1-1	1-2	1-3	2-1	2-2	2-3
平均順位	2.8	3.3	5.7	1.6	1.9	4.6

複があってもよいものとした。その結果、カメラ部と PC 部の持ちかえがない、コンパクトである等の理由から USB カメラを用いたシステムの方が高評価であった。DSC を用いたシステムを高評価とした被験者は画質やズーム等カメラとしての機能を重視していた。したがって、市販の DSC のようにカメラとして十分な機能を備えた上で、提案するキャプションングの機能を含むシステムが良いと考えられる。また、12 人中 5 人の被験者がシステム 1-1 と 1-2、2-1 と 2-2 を同じ順位という評価をしていた。これは後の被験者ほど地理情報データベースにデータが登録・更新された状態となり、データベースから提示される候補からの選択のみでキャプションを付加できることが多く 2 つのシステム間で作業に差がなかったためと考えられる。

4. ま と め

本稿では、写真を効率的に管理・共有することを目的とし、地理情報データベースと web 検索を用いた関連語抽出処理により、撮影位置・姿勢情報に基づくキャプションを半自動的に生成するシステムのプロトタイプを作成し、キャプション入力作

業に関する評価実験を行い、結果について考察した。提案システムは、地理情報データベースによってキャプションを提示できた場合、撮影からキャプション付加までを 10 秒程度で行うことができ、ユーザに負担をかけずにキャプションングを行うことができるものと考えられる。web 検索を用いた関連語抽出処理により候補を提示した場合、候補中の上位に選択するキャプションがある場合にはユーザの入力作業は簡単であったものの、下位にある場合には選択が煩雑になっていた。ユーザに提示されるキャプション候補が多い場合は、候補として提示されていても選択の作業は煩雑になる。これを解決するため、ユーザによるキャプションのキー入力に伴って提示されている候補を絞り込む機能等を検討している。また、ユーザの入力により新たなキャプションがデータベースに登録、またそのデータが更新される際に対応付けられる位置は、提案手法では実際の位置より撮影位置に近いものとなっており、より実際の位置に近づけるための手法の改良が必要であると考えられる。

その他の課題としては、建物が密集しているなど対象がより複雑である場合におけるシステムの挙動を調べる必要がある。特に地理情報データベース内でキャプション候補となり得るデータがより多くなる場合について、さらに実験を行う予定である。また、ユーザが入力するキャプションを単なる対象の名称ではない自由度の高いものにするための検討も必要である。

文 献

- [1] J. Electronics and I. T. I. Association (JEITA): "Exchangeable image file format for digital still cameras: Exif version 2.2" (2002).
- [2] Y. Wu, E. Y. Chang and B. L. Tseng: "Multimodal meta-data fusion using causal strength", Proc. 13th ACM annual Int. Conf. on Multimedia, pp. 872-881 (2005).
- [3] R. Sarvas, E. Herrarte, A. Wilhelm and M. Davis: "Metadata creation system for mobile images", Proc. 2nd Int. Conf. on Mobile Systems, Applications, and Services, pp. 36-48 (2004).
- [4] M. Naaman, R. B. Yeh, H. Garcia-Molina and A. Paepcke: "Leveraging context to resolve identity in photo albums", Proc. 5th ACM/IEEE-CS Joint Conf. on Digital Libraries, pp. 178-187 (2005).
- [5] 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).
- [6] 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).
- [7] M. Naaman, Y. J. Song, A. Paepcke and H. Garcia-Molina: "Automatic organization for digital photographs with geographic coordinates", Proc. 4th ACM/IEEE-CS Joint Conf. on Digital Libraries, pp. 53-62 (2004).
- [8] 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).
- [9] 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 (2005).
- [10] "Google Web API". <http://api.google.com/>.
- [11] 松本: "形態素解析システム「茶釜」", 情報処理, 41, 11, pp. 1208-1214 (2000).