

# ジオタグ付き写真を用いた写真の撮影方向の推定手法

廣田 雅春<sup>†</sup> 石川 博<sup>††</sup> 横山 昌平<sup>†††</sup>

<sup>†</sup> 静岡大学創造科学技術大学院/日本学術振興会特別研究員 DC ㊦ 432-8011 静岡県浜松市中区城北 3-5-1

<sup>††</sup> 首都大学東京システムデザイン学部情報情報通信システムコース ㊦ 191-0065 東京都日野市旭が丘 6-6

<sup>†††</sup> 静岡大学大学院情報学研究科, ㊦ 432-8011 静岡県浜松市中区城北 3-5-1

E-mail: <sup>†</sup>dgs11538@s.inf.shizuoka.ac.jp, <sup>††</sup>ishikawa-hiroshi@sd.tmu.ac.jp, <sup>†††</sup>tyokoyama@inf.shizuoka.ac.jp

あらまし 写真の撮影方向情報は、撮影者の関心を表す重要な情報としていくつかの研究で用いられている。しかし、その情報が付与された写真は未だに少ない。また、写真に撮影方向情報が付与されていた場合でも、誤った撮影方向を表している場合がある。そのため、正確な撮影方向情報が付与された写真の数を増加させることは、重要な課題である。撮影方向情報を推定する手法のひとつとして、建物を表すタグが付与されている写真に対して、その建物の地点を推定し、写真の撮影方向情報を推定する手法が提案されている。しかしながら、この手法では、その建物が写真の中央で撮影されていない場合、正確に写真の撮影方向情報を推定することはできない。そのため、本論文では、その手法で推定された撮影方向情報に対して、写真に撮影された建物の位置に基づいて、撮影方向情報を修正する手法を提案する。提案手法では、ある地点で撮影された写真とその周辺で撮影された写真に対して、領域分割により抽出された領域を比較することで、写真に含まれる建物の位置を推定する。画像共有サイトから取得した写真に対して提案手法を適用することで、その結果についての評価・考察を行った。

キーワード メタデータ推定, ジオタグ付き写真, Flickr

## 1. はじめに

近年, Flickr<sup>(注1)</sup> や, Panoramio<sup>(注2)</sup> などの画像共有サイトをはじめとして, Web を通じて多量の写真データを活用可能な場面が増加している。Web 上で共有されている写真の中には, GPS 搭載のスマートフォンや, デジタルカメラによって撮影位置情報が付与されているものも少なくない。ここで, 撮影方向情報とは, 撮影者が写真を撮影した方向を表すメタデータのことである。近年, それらのデバイスの普及に伴い, 撮影方向情報が付与された写真は, 増加傾向にある [1]。しかし, 撮影方向情報が付与可能なデバイスは, 未だに一部の機種に留まっている。実際に, Flickr において, キーワード”London”で検索して得られた撮影位置情報が付与された写真 1,687,731 枚において, 撮影方向情報が付与されていた写真は, 約 2.7%である 45,193 枚であった。また, 撮影方向情報が付与されていた場合でも, 撮影方向情報が誤った撮影方向を表している場合もある [2]。

撮影方向情報は, 撮影者がどのような写真を撮影したかを表す重要なメタデータとしていくつかの研究で用いられている。撮影方向情報が付与された写真を用いた研究の例として, 大規模な写真コレクションの閲覧支援 [3] や, 人々の興味のある領域の抽出 [4] [5] などがあげられる。これらの研究において, 撮影方向情報は, それらを算出するための特徴量として用いられている。そのため, 正確な撮影方向が付与された写真の数を増加させることは, 重要な課題である。

これまでに, 写真の撮影方向情報を推定するための手法がい

くつか提案されている [1], [6]。例えば, Zheng ら [1] の手法において, Google Street View<sup>(注3)</sup> や, Google Earth<sup>(注4)</sup> にマッチングできない写真や, 地面が写っていない写真に対しては, 撮影方向情報を推定することはできない。また, Google Street View への写真のマッチングを用いない手法として, 写真の被写体の位置を利用した手法も存在する。Huang ら [6] の手法では, ユーザーが指定した建物を表すタグが付与されている写真の撮影方向情報を推定する手法が提案されている。指定されたタグが付与されている写真から, そのタグの表す建物の中心位置を決定し, 写真の撮影方向をその写真の撮影位置から中心位置の方向を算出する。そのため, 写真の中央にそのタグが表す被写体が存在しない写真については誤差が生じる。図 1 に, 同一の被写体を撮影した横長な写真 (a) と縦長な写真 (b) を示す。これらの写真の被写体は, ロンドンにある Big Ben をほぼ同じ位置で撮影した写真である (ここで, Big Ben とは, ロンドンのウェストミンスター宮殿に付属する時計台のことである)。図 1(a) の写真において, 右側に Big Ben が位置するため, Huang らの手法を適用した場合, 実際の撮影方向と, 推定した撮影方向に誤差が生じる。

そこで, 本論文では, 写真の中央以外に被写体が位置する場合にも適切な撮影方向情報を推定可能な手法を提案する。これを実現するために, 撮影方向情報を推定したい写真に対して, 縦長の写真をマッチングすることで, 撮影方向の誤差を求める。図 1(a) のような横長の写真の場合は, 3 分割法などを考慮して写真の中心に主な被写体が存在しないことが多い。一方, 図

(注1) : <http://www.flickr.com/>

(注2) : <http://www.panoramio.com/>

(注3) : <http://www.google.co.jp/help/maps/streetview/>

(注4) : <http://www.google.co.jp/earth/>

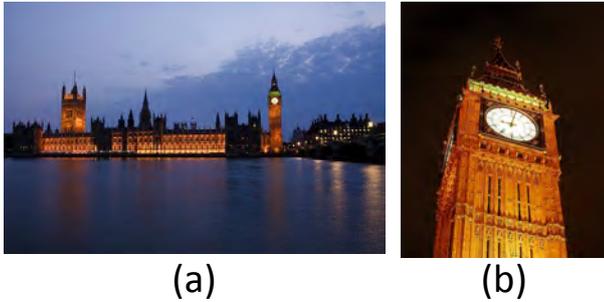


図 1 Big Ben を撮影したサイズの異なる写真

1(b) のような縦長の写真は、写真の主題となる被写体を写真の中心と撮影することが多い。また、横長の写真と比較して、縦長の写真の多くは、写真の大部分が主題となる被写体が領域を占めている。そのため、本論文では、縦長の写真から抽出した領域に基づいて、横長の写真にテンプレートマッチングすることで、被写体の画像の中心からのズレを算出する。提案手法では、はじめに、Huang らの手法を適用して、タグが表す被写体の中心位置を求め、撮影方向情報を推定する。その後、縦長の写真から特徴的な領域を抽出し、それぞれの写真から抽出された領域にマッチングすることで、タグが表す被写体の写真における位置を算出する。その結果とタグが表す被写体の中心位置の差を求めることで撮影方向情報を求める。

本論文の構成は、以下のようになっている。第 2 節では、撮影方向情報を用いた研究や、撮影方向情報を推定する関連研究について述べる。第 3 節では、Huang らの手法と、本論文で提案する被写体の位置に基づいた撮影方向情報の修正について述べる。第 4 節では、提案手法の評価と考察について述べる。第 5 節では、本論文で得られた成果をまとめ、今後の課題について述べる。

## 2. 関連研究

### 2.1 撮影方向メタデータの利用

写真の撮影方向情報を表すメタデータとして Exif がある。これはデジタルカメラなどで撮影された画像に付与されるメタデータである。記録されるメタデータとして、撮影方向や、撮影地点、撮影日時、カメラの機種名、焦点距離などがあげられる。これらの情報は、写真のクラスタリング [7] や、地理的特徴の抽出 [8]、イベント抽出 [9]、写真の撮影状況の分類 [10] などに用いられる。そのなかでも、近年、写真の撮影方向情報を利用した研究が増加している [1]。

写真の撮影方向情報は、写真のコレクションの閲覧支援や、写真によるランドマークなどの建物の 3D モデルの作成、写真の撮影位置の推定などに用いることが可能である [1]。その中でも、写真の撮影方向を用いて被写体の識別を行う研究が増加している。Shirai ら [4] は、人々の関心領域を可視化するために、写真の撮影方向に基づいて撮影スポットや、その撮影スポットが注目しているランドマークを関心領域として抽出する手法を提案している。他の撮影方向情報を利用した関心領域の研究として、Lacerda ら [5] の研究があげられる。Lacerda らは、複

数の写真から撮影方向の交点を算出し、それらを用いて関心領域を算出する。また、Epshtein ら [3] は、写真のコレクションの閲覧を様々な観点で実現するため、写真の撮影位置情報・撮影方向情報を用いて、建物の内側で撮影された写真と外側で撮影された写真を分類する手法を提案した。

これらの手法は、撮影方向情報が付与されている写真のみを対象としている。しかし、第 1 節で述べたように、Web 上で共有されている写真のほとんどは、撮影方向情報が付与されていない。そのため、撮影方向情報を特徴量として用いる手法を適用可能な写真が限定されてしまう。そのような手法において、撮影方向を示すメタデータの欠損による影響を少なくすることは、撮影方向情報を用いる研究の課題の 1 つである。

### 2.2 撮影方向情報の推定

写真の撮影方向情報を推定するための手法はいくつか提案されている [11], [12]。写真の撮影方向情報を推定するための手法として、撮影位置情報が付与された写真を路上のパノラマ画像に対してマッチングすることで撮影方向情報を推定する手法が提案されている。例えば、Zheng ら [1] の手法は、写真の撮影位置情報から取得可能な Google Street View の画像に対して写真をマッチングして撮影方向情報を推定する。マッチングに失敗した写真については、写真に写っている地面の領域と Google Earth をマッチングすることで、撮影方向情報を推定可能な写真の枚数を増やしている。また、Park ら [13] は、Google Street View に 2 つのフィルタを組み合わせて写真をマッチングすることで、撮影方向情報を推定する手法を提案している。1 つは、写真の撮影地点において、Google Street View が利用可能な場合は、それに対して画像をマッチングすることで撮影方向情報を推定する。もう 1 つは、写真の撮影地点の Google Satellite View が利用可能な場合は、写真に写っている地面に着目して撮影方向を推定する。丸山ら [14] は、画像の局所パターンに基づいて、路上から撮影した写真を Google Street View にマッチングすることで撮影方向を推定する手法を提案している。

これらの手法は、Google Street View などの車道から周辺を撮影した画像データを用いるため、車道付近で撮影された写真に対しては撮影方向が推定可能である。しかし、これらの手法には、複数の課題がある。1 つは、道路の周辺以外で撮影された写真に対して適用が困難なことである。公園の内部で撮影された写真など、車の侵入が難しい地点で撮影された写真をマッチングすることは困難である。最後に、人や車などの被写体が大半の領域を占める写真とのマッチングは困難なことである。

これらの課題に対して、Google Street View などのパノラマ画像に対するマッチングを行うことなく撮影方向情報を推定する手法として、Huang らによる研究があげられる [6]。Huang らはユーザが指定したランドマーク (目印になる建物) を表すタグが付与されている写真の撮影方向を推定する手法を提案している。指定したタグが付与されている写真から、そのランドマークの位置を決定し、そのランドマークの方向を写真の撮影方向とする。この手法は、ユーザが指定したランドマークに関連するタグが付与されている写真に対してのみ、撮影方向を推定することが可能である。



図2 提案手法による写真の撮影方向情報推定の手順

この手法では、前述したように、写真の中央にそのタグが表す被写体が存在しない写真については、推定される撮影方向情報に誤差が生じる。そのため、本論文では、この手法により推定された写真の撮影方向情報の誤差を修正することで、より誤差の少ない写真の撮影方向情報の推定手法を提案する。

### 3. 提案手法

本節では、ある地点で撮影された撮影位置情報が付与された写真の撮影方向情報を推定する手法について述べる。本論文において、Huangらの手法に基づいて、写真の撮影方向情報を推定し、その結果に対して画像の中心からの角度のズレを算出する。これにより、Huangらの手法によって推定された写真の撮影方向情報を修正する。

提案手法の概要を図2に示す。はじめに、撮影方向情報を推定する写真と、その写真に写っている建物を表すタグ(e.g., "Big Ben")を与える。次に、与えられたタグが付与されている写真をFlickrから収集する。取得した写真に対して、Huangらの手法を適用して、そのタグが表す建物の中心位置を算出する。取得した建物の中心位置に基づいて、写真の撮影方向を推定する。次に、写真の撮影方向情報を推定するために、撮影方向情報を推定する写真の撮影地点の周辺で撮影された写真の撮影方向情報を利用する。これは、ある写真の撮影地点の周辺で撮影された写真において、写真に付与されたタグが共通する場合、写真の撮影が類似する方向に対して行われたと考えられるためである。周辺で撮影された写真と、撮影方向情報を推定する写真に対して領域分割を適用し、写真から主要な領域を抽出する。抽出した領域に対してパターンマッチングを行い、写真において建物がどこに位置するかを算出することで、画像の中心からの角度を算出する。その結果に基づいて、Huangらの手法により提案した。撮影方向情報を修正し、最終的な撮影方向情報とする。

#### 3.1 建物の位置に基づいた撮影方向情報の推定

本論文では、建物を表すタグの位置に基づいて写真の撮影方向情報を推定するために、Huangらの手法を用いて建物の位置を推定する。

はじめに、ユーザが与えたタグが付与された写真をFlickrか

ら取得する。ここで取得するすべての写真には、撮影位置情報が付与されているとする。本論文では、写真に付与されているExifに含まれるGPSLatitude(緯度)とGPSLongitude(経度)を写真の撮影位置情報とする。

取得した写真に付与された緯度経度情報に基づいて、 $y$ 座標と $x$ 座標で表されるグリッドを作成する。以下の式を適用することで、写真に付与された緯度経度情報を2次元グリッドのセルの座標 $(x, y)$ に変換する。

$$y = M_{height} - \frac{(Lat - Lat_{min}) * M_{height}}{Lat_{max} - Lat_{min}} \quad (1)$$

$$x = M_{width} - \frac{(Lng - Lng_{min}) * M_{width}}{Lng_{max} - Lng_{min}} \quad (2)$$

ここで、 $Lat$ は、写真に付与された緯度(ExifのGPSLatitude)の値を表す。 $Lat_{max}$ ,  $Lat_{min}$ ,  $Lng_{max}$ , および  $Lng_{min}$  は、 $Lat$ と $Lng$ の最小値と最大値を表す。 $M_{height}$ ,  $M_{width}$ は、生成するグリッドの高さと幅をそれぞれを表す。本論文では、 $M_{height}$ ,  $M_{width}$ は、生成するグリッドの1つのセルが約10mになるように設定した。結果として、生成されたグリッドの1つのセルには、約10mの範囲で撮影された写真が含まれている。

生成されたグリッドに対して、ガウシアンカーネルを適用することでグリッドを畳み込む。これは、ノイズによる影響を低減するためである。適用するガウシアンカーネルは以下の式で表される。

$$G(y, x) = \frac{1}{2\pi\sigma} * \exp\left(\frac{(y - y_0)^2 + (x - x_0)^2}{2\pi^2}\right) \quad (3)$$

ここで、 $\sigma$ は、ガウシアンカーネルのスケールを表す。また、 $y_0$ と $x_0$ は、ガウシアンカーネルを適用する注目セルを表す。

ガウシアンカーネルを適用して、畳み込まれたグリッドから、タグの表す建物の領域を決定し、その中心座標を求める。それぞれのセルについて、 $G(y, x) > \alpha * \max(G)$ を満たすセルを求める。ここで、 $\max(G)$ は、すべての $G(y, x)$ の最大値を表す。 $\alpha$ は、セルがその建物の領域であることを許容するかを決定するパラメータである。本論文では、Huangらの手法でも用いられていた $\alpha = 0.6$ とする。

また、本論文では、Huangらの手法により領域を決定した後、ノイズとなる領域の除去を行う。これは、1人のユーザが撮影した写真に対してまとめて緯度経度情報を付与した場合に、1つの位置に大量の写真がアップロードされることがあり、その位置を含むセルが誤って建物の領域として判定されるためである。また、写真を撮影するデバイスによって、Exifに記録される緯度と経度の桁数が異なるため、丸め誤差が発生する(デバイスによって小数点2桁~6桁程度になっている)。そのような地点を含むセルに含まれる写真の枚数は非常に多いが、周囲のセルではほとんど写真が撮影されていない。そのため、抽出された領域の中で、あるセルに対して隣接する周囲8セルの写真の枚数が10枚以下の場合、そのセルはノイズとみなし、建

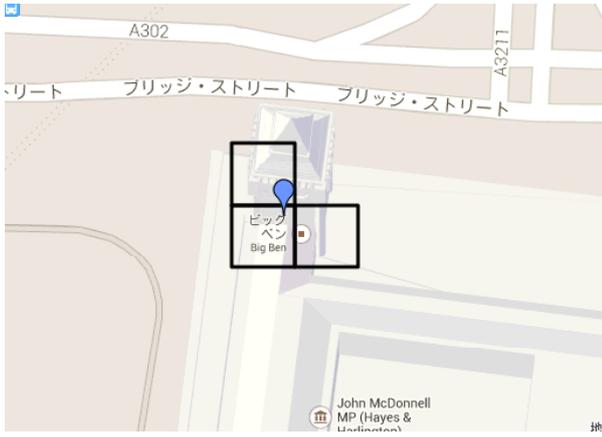


図3 タグ“Big Ben”が付与された写真から建物の中心を求めた結果

物の領域とはしない。求められた建物のセルの緯度経度の平均値を建物の中心位置とする。

図3に、タグ“Big Ben”が付与された写真から建物の中心位置を推定した結果を地図上に表示したものを示す。図3において、青いマーカーが算出されたBig Benの中心位置である。また、黒い四角形が建物の領域であると推定されたセルである。図3において、Huangらの手法により推定されたBig Benの中心位置と、実際のBig Benの位置の誤差は数mであり、ほとんど差がないことがわかる。

得られた建物の中心の緯度経度情報に基づいて、写真の撮影方向情報 $\theta_l$ を推定する。それぞれの写真の撮影位置から、建物の中心を撮影したと仮定して、以下の式で求める。

$$n = \cos(Lng) * \sin(Lat - Lat_l) \quad (4)$$

$$d = \cos(Lng_l) * \sin(Lng) - \sin(Lng_l) * \cos(Lng) * \cos(Lat - Lat_l) \quad (5)$$

$$\theta_l = \frac{\arctan(n, d) * 180}{\pi} \quad (6)$$

ここで、 $Lat_l$ と $Lng_l$ は、得られた建物の領域の中心の緯度経度である。

### 3.2 写真内での建物の位置に基づいた撮影方向情報の修正

本節では、建物が写真のどの位置で撮影されているかに基づいて、前節で推定した撮影方向情報の修正について述べる。

はじめに、撮影方向を推定する写真の撮影位置から $\gamma$ m以内で撮影された写真を取得する。

取得した写真と撮影方向情報を修正したい写真に、それぞれに前処理を行う。本論文で行う前処理は、画像の平坦化と、平滑化である。

平坦化は、写真を撮影した状況によって、天候や光のあたり方による色の変化や、色調の違いに対応するために適用する。画像をHSV空間で表し、その輝度であるYに対して平坦化を行い、再度RGBに変換する。

平滑化には、画像に含まれるのノイズを除去するためにメディアンフィルタを用いた。メディアンフィルタは、以下の式で表されるように、注目画素の近傍から求めた画素の中央値を注目がその新たな画素とする処理であり、画質を損なわずに平滑化することが可能である。

$$g(x, y) = \text{median}\{f(x + m, y + n) | -1 \leq m, n \leq 1\} \quad (7)$$

次に、減色処理を用いて画像の領域を分割する。減色処理は、RGB空間で表現された写真に対してk-means法を適用する。k-means法のアルゴリズムを以下に示す。

- (1) 色空間に画像の画素をすべて配置する。
- (2) クラスタの中心をランダムにK個配置する。
- (3) それぞれの画素をK個のクラスタに割り当てる。
- (4) クラスタの中心を再計算する。
- (5) 新しいクラスタが前のクラスタと十分に近ければ処理を終了する。さもなければ(3)に戻る。

しかし、それぞれの画像において、適切なクラスタ数が異なるため、本論文では、k-means法で用いるクラスタ数を多めにしておき、抽出されたそれぞれのクラスタについて以下の式を満たすクラスタ同士を結合する。本論文では、色相や明度の類似性に基づいてクラスタ同士を結合する。そのため、クラスタの重心 $(a, b)$ をHSV空間に変換した値に基づいて、以下のいずれかの条件を満たした場合に結合する。

$$|H_a - H_b| > 36$$

$$V_a, V_b \leq 25 \quad (8)$$

$$V_a, V_b \geq 230$$

ここで、 $H_a, H_b$ は、クラスタの重心 $(a, b)$ の色相 $H(0 \leq H \leq 360)$ を表す。また、 $V_a, V_b$ は、クラスタの重心 $(a, b)$ の明度 $V(0 \leq V \leq 255)$ を表す。1つ目の条件を満たした場合は、2つのクラスタの重心となっている色の色相が類似していることを表している。残りの2つの条件は、2つのクラスタの重心となっている色が極端に明るい場合と暗い場合を表している。

次に、ラベリングを行うことで、1つのクラスタに属する画素に同一のラベルを付与することで領域分割を行う。写真に含まれる1つのクラスタの領域が2つ以上存在した場合は、領域の範囲が大きい領域のみを取得する。また、極端に小さい領域についても、取得しない。これは、多くの場合、建物などではなく、光源などのあまり重要ではない領域であると考えられるためである。

次に、取得した領域同士をマッチングすることにより、建物が写真のどの位置で撮影されたかを算出する。本論文では、縦長な写真から取得した領域を撮影方向情報を推定したい写真から取得した領域にマッチングすることで行う。

縦長の写真から取得したそれぞれの領域と、撮影方向情報を推定したい写真から取得したそれぞれの領域をマッチングする。本論文では、一般的なZNCC(Zero-mean Normalized Cross Correlation)などによって計算される輝度を用いたテンプレートマッチングではなく、領域の重なりのみに着目してテンプレートマッチングを行う。これは、本論文において、写真

が撮影された時の照明などの明るさや、昼夜などの撮影状況によって、画像の色が大きく変化するため、形状にのみ基づいて、マッチングを行うためである。また、領域をマッチングする際に、それぞれの領域によって、サイズが異なるため、縦長の写真を撮影方向情報を推定したい写真の縦のサイズに基づいてリサイズを行う。テンプレートマッチングは以下の式を用いて行う。

$$R(y, x) = \frac{1}{|g|} \frac{1}{|f|} \sum \sum \begin{cases} 1 & g(y+i, x+j), f(i, j) = 1 \\ 0 & otherwise \end{cases} \quad (9)$$

ここで、 $g(i, j)$  は、撮影方向を推定したい写真から取得した領域の座標  $(i, j)$  を表す。 $f(i, j)$  は縦長の写真から取得した領域の座標  $(i, j)$  を表す。 $g(y+i, x+j) = 1, f(i, j) = 1$  は、それぞれの領域において、ラベリングされている領域を表す。 $|g|$  と  $|f|$  は、それぞれの領域において、ラベルが付与されている座標  $(f(i, j) = 1)$  の数を表す。すべての縦長の写真から取得したそれぞれの領域について、 $R_{(y,x)} > 0.6$  を満たす座標の平均値  $(y_t, x_t)$  をその写真の建物の位置とする。

最終的に推定される撮影方向情報  $\theta$  は、Huang らの手法で求めた建物の中心位置の方向  $\theta_l$  に対して、推定された建物の位置の写真の中心からのズレを用いて求める。建物の位置の中心からのズレは、写真の画角を用いた以下の式で求められる。

$$\theta = \theta_l + \frac{2 \arctan \frac{l}{2f}}{|x|} (x_t - \frac{|x|}{2}) \quad (10)$$

ここで、 $|x|$  は写真の横幅、 $f$  は写真の焦点距離、 $l$  は映像素子のサイズを表す。焦点距離は、写真に付与された Focal Length を用いる。

#### 4. 実 験

本論文では、Flickr から取得した写真に対して提案手法を適用し、その評価を行う。提案手法による撮影方向情報の推定と Huang らの手法による撮影方向情報の推定結果をそれぞれ求め、その結果を人手により判断された撮影方向との誤差を比較することで提案手法を評価する。

評価に用いる写真は、以下の条件をすべて満たす写真 26 枚である。

- (1) 写真に付与されている Exif の中に Focal Length を含む
  - (2) 写真に撮影位置情報が付与されている
  - (3) 指定したタグが写真に付与されている
  - (4) 指定したタグに関する被写体が写真に撮影されている
- Focal Length は、写真の画角を求めるために用いる。また、Huang らの手法において、1つの建物を表すタグを指定して、そのタグが付与されている大量の写真から、タグの表す建物の位置を推定しているため、写真の撮影位置情報とタグが写真に付与されている必要がある。本論文では、タグ”Big Ben”が付与されている写真 3,189 枚を用いた。また、写真に付与されて

表 1 提案手法と比較手法の推定誤差

	平均誤差
Huang らの手法	10.21
提案手法	6.44

いるタグと写真の意味するところが異なる場合があるため、本研究では、写真のなかに Big Ben が撮影されているもののみを評価に用いた。

これらの写真を人手により評価するために、Google Maps<sup>(注5)</sup> と、Google Street View を用いて、撮影地点の周辺の領域を比較することで、写真の撮影方向を決定した。

ここで、Huang らの手法によって推定された撮影方向情報と、提案手法によって推定された撮影方向情報のそれぞれを人手により判断された撮影方向との誤差の平均値を表 1 に示す。表 1 において、Huang らの手法と比較して、提案手法により推定された撮影方向情報の平均誤差は約 3.8 度減少している。そのため、提案手法は、写真に含まれる建物の位置を算出し、撮影方向情報を推定できていることがわかる。図 4 に、提案手法によって撮影方向情報が推定された写真の例を示す。4 枚の写真は、それぞれ”Big Ben”を撮影した写真である。写真に記されている青い丸点は、提案手法により推定された”Big Ben”の位置である。図 4 より、実際の”Big Ben”と提案手法により推定された位置が一致しており、提案手法が有効に機能していることがわかる。これにより、提案手法を適用することで、実際の建物の位置と写真の中での建物の位置の関係がわかるため、Huang らの手法による撮影方向情報の推定に対して、写真の中での建物の位置が中心でなかった場合についても、適切に撮影方向を推定できると考えられる。

図 5 に、提案手法を用いて推定した撮影方向情報と、人手により判断された撮影方向との誤差の累積分布を示す。図 5 において、約 80% の写真に対して、提案手法により推定された撮影方向情報の誤差は、8 度以下である。そのため、多くの写真に対して、提案手法が効果的に機能していることがわかる。しかし、一部の写真については、人手により判断された撮影方向と 20 度以上の誤差がある。これについては、提案手法の領域分割が適切に機能しなかったことが原因としてあげられる。図 6 に、写真とその写真を領域分割した結果を示す。図 6 において、青い点が提案手法によって推定された撮影方向情報である。また、紫の枠で囲まれている画像は、その中の白い部分が抽出された領域である。領域分割を行った際に、”Big Ben”と雲がひとつの領域として抽出されてしまっている(右下の領域)。そのため、周辺で撮影された写真とマッチングした際に、形状が類似していないと判断されたと考えられる。

以上より、提案手法が効果的に機能する場合は、写真の領域分割の性能に依存する。しかし、多くの写真では、提案手法を用いることで、撮影方向情報の推定について結果の改善が見られた。

(注5) : <https://maps.google.com/>



図4 提案手法により撮影方向情報が推定された写真の例

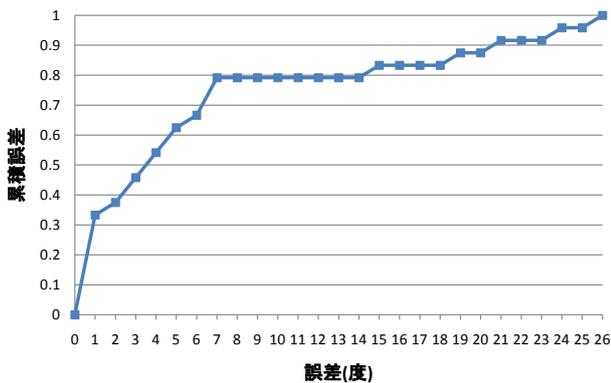


図5 提案手法を用いて推定した撮影方向情報の誤差の累積分布

## 5. 終わりに

本論文では、写真に撮影された建物の位置に基づいて、撮影方向情報を修正する手法を提案した。タグと撮影位置情報が付与された写真の撮影方向情報を推定するため、はじめに、Flickr にアップロードされているそのタグが付与された写真から、そのタグの表す建物の位置を推定し、その位置に向かって写真は撮影されたとする。次に、その結果に対して、周辺で撮影された写真とのマッチングを行い、建物が写真のどの位置で撮影されたかを算出することにより、写真の撮影方向情報を推定する。また、提案手法により推定された撮影方向情報と人手により判断された撮影方向情報と比較することによって、提案手法の評価を行なった。提案手法を適用することで、多くの写真に対して建物が写真のどの位置で撮影されたかを抽出することができ、Huang らの手法によって推定された撮影方向情報に含まれる誤差の低減に提案手法が有効であることを確認した。

今後の課題として、以下の2つがあげられる。1つは、提案手法における、画像の領域分割の精度の向上があげられる。前述したように、提案手法は、画像の領域分割の性能に大きく依存する。そのため、本論文のように、RGB 値に基づいた領域分割以外の手法についても検討する必要があると考えられる。もうひとつは、テンプレートマッチングする領域の選出があげられる。撮影方向情報を推定したい写真から抽出した領域と、

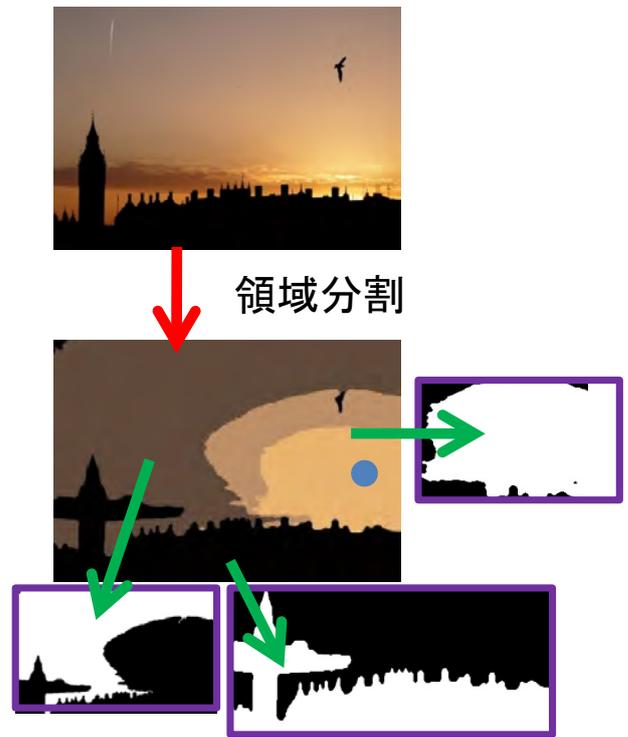


図6 提案手法による撮影方向情報の推定に失敗した写真とその領域分割の結果の例

縦長の写真から抽出した領域をマッチングしているが、実際にマッチングすべき領域は、建物の領域のみである。そのため、抽出した領域に対して、教師あり学習を適用し、建物か、それ以外の領域 (e.g., 空, 地面) などについては、テンプレートマッチングを行わないことにより、建物の位置の画像の中心からのズレの算出の性能の向上が期待される。

## 文 献

- [1] Y.-T. Zheng, Z.-J. Zha and T.-S. Chua: "Research and applications on georeferenced multimedia: a survey", *Multimedia Tools Applications Archive*, **51**, 1, pp. 77-98 (2011).
- [2] R. B. Smith, B. G. Morton and M. R. Elgersma: "Electronic compass and compensation of large magnetic errors for operation over all orientations", *US Patent* (2003).
- [3] B. Epshtein, E. Ofek, Y. Wexler and P. Zhang: "Hierarchical photo organization using geo-relevance", *Proc. 15th annual ACM international symposium on Advances in geographic information systems*, Seattle, USA, pp. 18:1-18:7 (2007).
- [4] M. Shirai, M. Hirota, S. Yokoyama, N. Fukuta and H. Ishikawa: "Discovering multiple hotspots using geo-tagged photographs", *Proc. 20th International Conference on Advances in Geographic Information Systems*, California, USA, pp. 490-493 (2012).
- [5] Y. A. Lacerda, R. G. F. Feitosa, G. A. R. M. Esmeraldo, C. d. S. Baptista and L. B. Marinho: "Compass clustering: a new clustering method for detection of points of interest using personal collections of georeferenced and oriented photographs", *Proceedings of the 18th Brazilian symposium on Multimedia and the web*, WebMedia '12, New York, NY, USA, ACM, pp. 281-288 (2012).
- [6] Y.-T. Huang, K.-T. Chen, L.-C. Hsieh, W. Hsu and Y.-F. Su: "Detecting the directions of viewing landmarks for recommendation by large-scale user-contributed photos", *Proc.*

- 20th ACM international conference on Multimedia, Nara, Japan, pp. 997–1000 (2012).
- [7] M. Hirota, S. Yokoyama, N. Fukuta and H. Ishikawa: “Constraint-Based Clustering of Image Search Results Using Photo Metadata and Low-Level Image Features”, *Computer and Information Science 2010*, pp. 165–178 (2010).
  - [8] C. Sengstock and M. Gertz: “Latent geographic feature extraction from social media”, *Proceedings of the 20th International Conference on Advances in Geographic Information Systems, SIGSPATIAL '12*, New York, NY, USA, ACM, pp. 149–158 (2012).
  - [9] L. Chen and A. Roy: “Event detection from flickr data through wavelet-based spatial analysis”, *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*, New York, NY, USA, ACM, pp. 523–532 (2009).
  - [10] P. Sinha and R. Jain: “Classification and annotation of digital photos using optical context data”, *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, CIVR '08*, New York, NY, USA, ACM, pp. 309–318 (2008).
  - [11] X. Xu, T. Mei, W. Zeng, N. Yu and J. Luo: “Amigo: Accurate mobile image geotagging”, *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, ICIMCS '12*, ACM, pp. 11–14 (2012).
  - [12] F. X. Yu, R. Ji and S.-F. Chang: “Active query sensing for mobile location search”, *Proceedings of the 19th ACM International Conference on Multimedia, MM '11*, ACM, pp. 3–12 (2011).
  - [13] M. Park, J. Luo, R. T. Collins and Y. Liu: “Beyond gps: determining the camera viewing direction of a geotagged image”, *Proc. international conference on Multimedia, Firenze, Italy*, pp. 631–634 (2010).
  - [14] 丸山拓馬, 柳井啓司: “位置情報付き路上画像の撮影方向推定システムの提案”, *情報学 CVIM 研報*, **2010**, 12, pp. 1–8 (2010).