

Geometric Context を用いた Randomized Trees による画像撮影位置同定

福田 考晃[†] 山内 一祥[†] 藤吉 弘亘[†]

[†] 中部大学大学院工学研究科 〒487-8501 愛知県春日井市松本町 1200

E-mail: [†]{fukuta,ky}@vision.cs.chubu.ac.jp, ^{††}hf@cs.chubu.ac.jp

あらまし 本稿では、画像の撮影位置を高速に同定する手法を提案する。従来の撮影位置同定手法は、参照用画像との特徴点マッチングや、テンプレートマッチング結果を元に撮影位置を同定する。そのため、同定までの処理コストが高いという問題や、季節や天候の変化による見えの変化が同定精度に影響を与えるという問題がある。そこで、季節の変化を受けにくい Geometric Context を用いた特徴量と、統計的学習手法である Randomized Trees を用いて撮影位置を同定する。これにより、従来のマッチング結果に基づいた手法を用いた場合と同等の精度でかつ速度を 10^3 倍向上させることができた。

キーワード 撮影位置同定, Geometric Context, ランダム木

Image Based Localization with Randomized Trees Using Geometric Context

Takaaki FUKUTA[†], Katsuyoshi YAMAUCHI[†], and Hironobu FUJIYOSHI[†]

[†] Dept. of Computer Science, Chubu Univ. 1200, Matsumoto, Kasugai, Aichi, 487-8501 Japan

E-mail: [†]{fukuta,ky}@vision.cs.chubu.ac.jp, ^{††}hf@cs.chubu.ac.jp

Abstract This paper proposes an algorithm for classifying image camera positions with high speed. Conventional camera position classification algorithms classify camera positions by matching feature points against a reference image or using template matching results. The drawbacks of these conventional algorithms are high processing costs for the processes through classification, and variable classification precision due to the effect of changing seasons or weather on view transformations. The proposed algorithm classifies camera positions by using feature values that employ a geometric context not easily affected by weather changes, and the Randomized Trees statistical learning algorithm. In comparison to a conventional algorithm based on matching results, the proposed algorithm provides the same precision while increasing speed by a factor of 10^3 .

Key words Image Based Localization, Geometric Context, Randomized Trees

1. はじめに

高精度に現在位置を同定する技術は、ナビゲーションシステムや自立型ロボットへの応用が期待され、重要な技術である。ナビゲーションシステムでは、高精度化に伴い自動制御などの付加価値を利用したナビゲーションをすることが可能となる。また、自律型ロボットでは、現在位置を高精度に得られることで、自律行動範囲の拡大が期待できる。

位置情報が取得できる機器として GPS がある。GPS は携帯電話に搭載されるなど一般化が進み、精度が向上している。しかし、携帯電話などに搭載されている GPS は高層ビルなどの遮蔽物が多い場所など、条件によっては正しく位置情報が得られない場合がある [8]。さらに、GPS で得られるのは位置情報であり、方位を求めること

ができない。

この問題を画像を用いて解決する手法が提案されており、それらは特徴点マッチングに基づいた手法とシーンのマッチングに基づいた手法に分けることができる。特徴点マッチングに基づいた手法は、SIFT [10] に基づいた手法がそのほとんどである [4] [5] [12] [13] [16]。特徴点マッチングを用いることで、都市で撮影された画像に対し、高い同定精度を得られることが報告されている。しかしながら、特徴点の投票を Nearest Neighbor により行い、画像参照を行うため、探索時間が膨大である問題や、木など複雑な見えを含むシーンからは特徴点が無数に算出されてしまい、同定精度を低下させる問題がある。

シーンのマッチングに基づいた手法では、様々なシーンを記述することができる特徴量を用いてマッチングを行う IM2GPS [3] が提案されている。この手法は、Flickr [9]



図 1 同位置における季節と天候が異なる例

からダウンロードした撮影条件が未知の画像を対象としており、人工物や、特徴点ベースの手法において困難となる自然画像に対応した例である。色情報や画像中の線などを利用した様々な特徴量を用いているため、海や森、またランドマークとなる建造物など、シーンに依存することなく撮影位置を同定することが可能である。しかしながら、色情報を用いているため、図 1 のように、季節や天候の変化により見えが変化したものには対応できないという問題がある。また、特徴点マッチングに基づいた手法と同様に Nearest Neighbor を用いているため、探索時間が膨大となり、結果位置同定にかかる時間が増加するという問題がある。

従来法の共通の問題である探索時間の増加は、リアルタイム性が必要なナビゲーションシステムやロボットビジョンにおいては大きな問題となる。さらに、季節の変化に対応することができなければ、各季節の参照用画像を用意する必要がある、さらなる探索時間の増加や画像収集にかかる手間が膨大となる問題を招く。

そこで、本稿では、季節の変化に影響を受けにくい Geometric Context [11] を用いた特徴記述と統計的学習手法である Randomized Trees [1] を用いることで、処理コストの低い撮影位置同定手法を提案する。

本稿では、2. 章で関連研究、3. 章で提案手法について述べる。そして、4. 章で提案手法の評価実験を行い、5. 章でまとめを行う。

2. 関連研究

画像を用いた位置同定手法は、特徴点マッチングを用いたものと、シーンマッチングを用いたものに分けられる。特徴点マッチングを用いた手法として、Cipora らは都市の画像を対象に建物領域をラベリングしたデータを用いて、建物の正面化を行った後にマッチングを行うことで、幾何学的変化に対応した手法を提案している [13]。同様に上野らは建物の正面化を行うが、画像中の領域を Bag of Features [8] により分類することで建物領域のラベリングを必要とせずに建物の正面化を行い、その後 SIFT 特徴量を用いて画像間のマッチングを行う手法を提案している [5]。また、ICCV'05 の “Where am I ?” コンテストで最も良い結果を残した Zhang らは SIFT によるマッチングの枠組みで投票処理を行う際に、通常のユークリッド距離だけでなく、コサイン距離を利用することで幾何平面を推定し高精度に撮影位置を同定する手

法を提案している [16]。小野らは独自に構築したデータベースを用い、近藤らにより提案されたブロックマッチング法 [6]、SURF [17] の特徴点を用いた投票により類似度を算出することで、撮影位置同定の検討を報告している [4]。最後に、薄らはあらかじめ構築したランドマークデータベースを用いて、静止画像からの撮影位置同定手法を提案している [12]。

特徴点マッチングを用いる際の問題点は 2 点あり、一つ目は、特徴点の投票に時間がかかるという点である。二つ目は、建物から算出される特徴点は類似したものが多いことや [13]、木などの自然物が存在する複雑なシーンからは、特徴点が大量に算出されてしまうという点である。

そこで、特徴点マッチングを用いないシーンマッチングを用いた手法が提案されている。Hays らにより提案された IM2GPS [3] は、ここまでの従来法とは異なり、対象を Flickr からダウンロードした撮影条件の制約がない画像データとし、撮影位置同定を検討しており、シーン記述に用いられる特徴量を数種類用いることで画像マッチングを行い、撮影位置を同定する。彼らが用いた特徴量は次の 6 つである (1) Tiny Images, (2) カラーヒストグラム, (3) Texton ヒストグラム, (4) Line Features, (5) 色情報を用いた Gist 記述子, (6) Geometric Context [11]。これらすべての特徴量を用いた Nearest Neighbor 法による最近傍画像探索により撮影位置を同定する手法を提案しており、実験により様々なシーンに対応することが可能であることを示している。しかしながら、色情報を用いていることにより季節や天候の変化に対応できないことや特徴点マッチングを用いた手法と同様に Nearest Neighbor を用いているため、探索時間が膨大であるという問題がある。

3. 提案手法

提案手法は、季節や天候の変化を受けにくい Geometric Context と識別処理が高速である Randomized Trees を用いることで撮影位置の同定を実現する。

提案手法を図 2 に沿って説明する。まず、画像から Geometric Context の確信度画像を算出し、それらから [2] で用いられている特徴量を Geometric Context の確信度画像に拡張し、共起特徴を抽出する。これを Randomized Trees により識別することで、類似画像を探索する。探索された類似画像の位置情報を用いることで入力画像の撮影位置を求めることが可能である。

3.1 Geometric Context

Hoiem らにより提案された Geometric Context (GC) は、一枚の入力画像に対して三次元の面構造を推定し、ラベリングする手法である。この手法は、色 (16 特徴量) やテクスチャ (15 特徴量)、位置 (8 特徴量)、形状 (4 特徴量)、三次元構造 (35 特徴量) の 73 特徴量を入力とし、

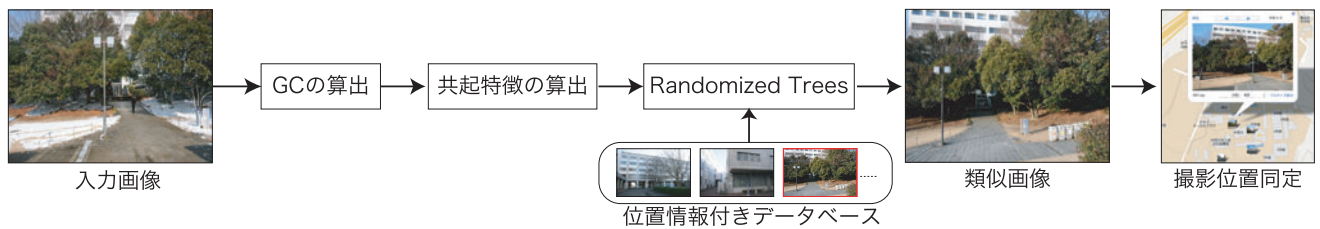


図2 提案手法の流れ

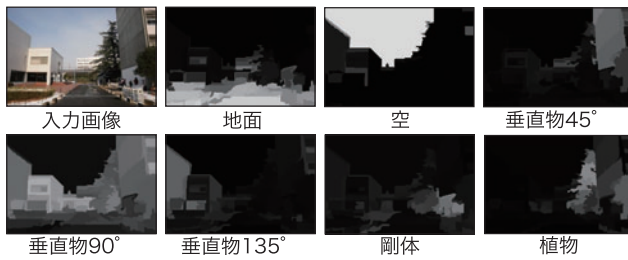


図3 Geometric Context

ロジスティック回帰を利用した AdaBoost によりモデルを学習している。識別時には、学習時に得られたモデルを用いて、未知入力画像に対して「地面 (ground)」、「垂直物 (vertical)」、「空 (sky)」に分類する (図3)。ここで、垂直物は 45° 、 90° 、 135° 、剛体、植物に分けることができ、提案手法ではこれらを個別に扱う。学習は、まず入力画像を Felzenswalb らが提案した手法 [18] により super pixel と呼ばれる領域に分割する。次に、分割された領域から色テクスチャ、位置、形状、三次元構造の特徴量を抽出し、弱識別器に (1) 式を用いた AdaBoost により、幾何学的に均一な領域となるように最適化を行う。

$$f_m(x_1, x_2) = \sum_i^{n_f} \log \frac{P(y_1 = y_2, |x_{1i} - x_{2i}|)}{P(y_1 \neq y_2, |x_{1i} - x_{2i}|)} \quad (1)$$

ここで、 x_i, x_2 は隣接する super pixel の持つすべての特徴量、 y_1, y_2 は正解ラベル、 n_f は特徴数である。

最後に、次式を用いてセグメンテーションされた領域に「地面」、「垂直物 $\times 5$ 」、「空」の7つのラベルの確信度 C を求める。

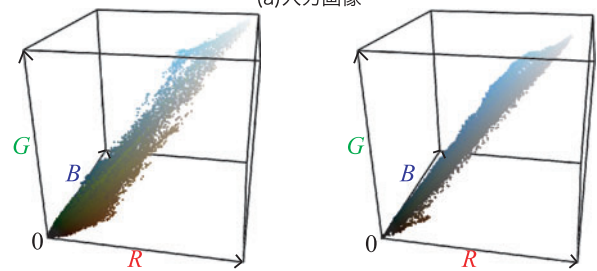
$$C(y_i = e|x) = \sum_j^{n_h} P(y_j = e|x, h_{ji})P(h_{ji}|x) \quad (2)$$

ここで、 y はクラスのパベル、 e は推定ラベル、 x は画像データ、 n_h は領域分割された数、 h_{ji} は AdaBoost により最適化された領域に含まれる super pixel の番号である。右辺の確率尤度関数はそれぞれ、第一項が「領域が対象ラベルである確信度」、第二項が「領域の均一性」を表現している。図4に季節、天候が異なる2枚の画像に GC を用いた例を示す。(b) の RGB 色空間では11月に撮影したものの R 成分が強く、1月撮影にはほとんど見られないことが分かる。これにより、色情報を用いてこれら2枚の画像同様のシーンであることを示すことが難しいといえる。しかし、GC の確信度画像(c)(d)(e)を見てみると、地面や空、垂直物を色の変化に対して頑健にラベリング出来ていることが分かる。これは、GC では色以外にテクスチャ、位置、形

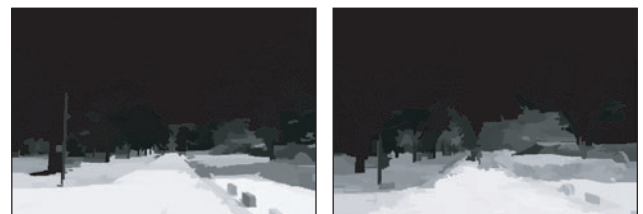
状、三次元構造を用いているため色のみには依存しないためであると考えられる。これにより、GC を用いることにより、季節や天候の変化を受けること無い特徴抽出が期待できる。



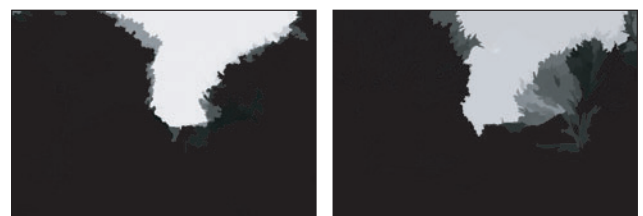
(a)入力画像



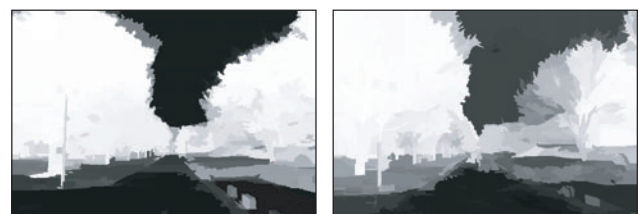
(b)RGB色空間の分布



(c)地面



(d)空



(e)垂直物

図4 季節天候が異なる画像から得られる GC

3.2 GC 間の共起を用いた特徴表現

提案手法では、季節や天候の変化に頑健な特徴記述を

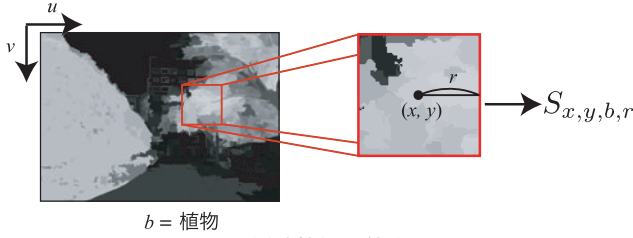


図5 領域特徴の算出方法

行うために、GCを用いて特徴記述を行う。[3]では、GCにより得られる画像をダウンサンプリングし、大域的な情報のみを用いて画像マッチングを行っているが、ダウンサンプリングにより細部の情報が失われてしまう問題がある。この問題を避け、頑健にシーンを記述するために局所領域間の共起を利用した特徴量を用いる。

GCにより得られる確信度画像7枚を用いて、特徴量を算出する領域を2つ用いることにより共起を表現する。算出領域は図5に示すように、注目画像座標を x, y 、確信度画像のインデックス $B = \{b \mid \text{垂直物 } 45^\circ, \text{垂直物 } 90^\circ, \text{垂直物 } 135^\circ, \text{地面, 空, 剛体, 植物}\}$ 、領域サイズを r とし、 $p_{x,y,b}$ をGCの確信度とすることにより、以下の式により領域のGC特徴を算出する。

$$S_{x,y,b,r} = \sum_{j=x-r}^{x+r} \sum_{i=y-r}^{y+r} p_{i,j,b} \quad (3)$$

上式により求めることができる S を、[2]で用いられている特徴量と同様に以下の4つの数式を用いることにより共起を表現する。

$$f = S_{x,y,r,b} \quad (4)$$

$$f_{sum} = S_{x_1,y_1,r_1,b_1} + S_{x_2,y_2,r_2,b_2} \quad (5)$$

$$f_{diff} = S_{x_1,y_1,r_1,b_1} - S_{x_2,y_2,r_2,b_2} \quad (6)$$

$$f_{abs} = |S_{x_1,y_1,r_1,b_1} - S_{x_2,y_2,r_2,b_2}| \quad (7)$$

f はある確信度画像 b のある領域 $S_{x,y,b,r}$ の値を特徴量とするものであり、 $f_{sum}, f_{diff}, f_{abs}$ はある確信度画像 b_1, b_2 のある領域 S_{x_1,y_1,b_1,r_1} と S_{x_2,y_2,b_2,r_2} の関係を捉える特徴量である。これらの特徴量を用いて、Randomized Treesを学習することで、高速な撮影位置同定を行う。

3.3 Randomized Trees

Randomized Trees は、木構造を持つフレームワークであるため、識別が高速であるというメリットがあり、実用的な問題にも利用され始めている。Randomized Trees は、識別、回帰、クラスタリングなどに用いることができることが知られている。提案手法では、撮影位置の識別に Randomized Trees を用い、撮影位置をあらかじめ関連づけた画像を用いて Randomized Trees を構築する。これにより、高速に類似画像を探索することが可能となり、従来法の問題であった膨大な計算コストを軽減する。識別に用いられる Randomized Trees は大きく分けて2つの派生を持ち、それらは分岐を評価する方法が

異なる。一方は情報利得を用いることで分岐を評価し、Randomized Trees を構築する方法である [2]。他方は、情報利得を用いず、ランダムに構築する方法である [7]。前者は、分岐を情報利得により評価を行うため、学習サンプル数に依存しないが、学習時間は長くなる。後者は、分岐を評価することを行わないため、学習サンプル数が十分多い場合でないと上手く学習できないが、学習時間は短くて済む。本稿の実験において、評価に用いるデータベースは比較的画像枚数が少ないため、前者の情報利得を利用するものを用いる。

3.3.1 学習

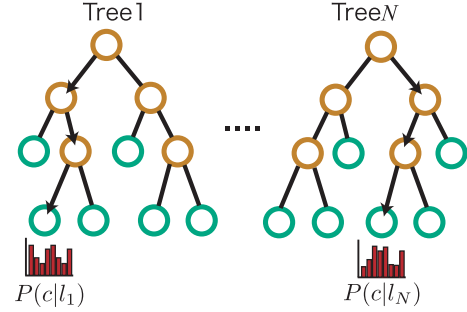


図6 Randomized Trees(黄丸：分岐ノード，緑丸：末端ノード，末端ノードが持つカテゴリ分布 $P(c|l)$)

Randomized Trees は、複数のランダムな要素を取り入れて構築された決定木の集合である (図6)。つまり、複数の決定木を構築することが Randomized Trees の学習にあたる。複数の決定木を構築する為に、教師信号 (提案手法では撮影位置) 付き入力データ I をサブセットと呼ばれるデータ集合 $I' \subseteq I$ にわけ、そして、サブセットひとつに対し、ひとつの決定木を構築する。サブセットを利用することにより、過学習の抑制や、学習の効率化といった効果が得られる。各決定木の構築は、入力データを分岐ノードにより分岐することで行う。各分岐ノードに対し、入力データが I_n 、左右に分岐するデータがそれぞれ、 I_l, I_r であったとすると、分岐条件は (8) 式、(9) 式で表される。

$$I_l = \{i \in I_n | f(v_i) < t\} \quad (8)$$

$$I_r = I_n \setminus I_l \quad (9)$$

$f(v_i)$ は特徴量、 t はしきい値である。ここで、 $f(v_i)$ と t はランダムに選択する。上記のデータの分岐をさせた後、(10) 式で情報利得 (Information gain) ΔE を求め、その分岐を評価する。

$$\Delta E = -\frac{|I_l|}{|I_n|} E(I_l) - \frac{|I_r|}{|I_n|} E(I_r) \quad (10)$$

ここで、 $E(I)$ とは、情報エントロピーであり、データサンプルに付与されている教師データを用いて次のように算出する。 $E(I) = -\sum_{i=1}^n P_i \log_2 P_i$ 、 P_i はデータ集合 I に含まれる教師データから得られるカテゴリの分布、 n は学習カテゴリの数である。以上の処理を繰り返すことで、様々な特徴量としきい値の組み合わせをランダムで

選択し、最も (10) 式が大きくなった組み合わせをその分岐ノードのパラメータと決定する。構築の終了条件は、分岐ノードの深さが設定した値までの深さになるか情報利得が限りなく小さくなるまでとする。末端ノード l には、分岐したデータ I_n の教師信号によりカテゴリ分布 $P(c|l)$ が作成される。

3.3.2 識別

Randomized Trees で識別をするためには、学習により作成された複数の決定木の末端ノード l のカテゴリ分布 $P(c|l)$ を用いる。複数の木をトラバースルすることにより、たどり着いた末端ノードの分布の平均を以下の式で求める。

$$P(c|L) = \frac{1}{T} \sum_{t=1}^T P_t(c|l_t) \quad (11)$$

T は決定木の数、 $L = (l_1, \dots, l_T)$ である。この式により、画像を入力した時に c 地点である事後確率を求めることができる。そのため、入力画像のクラスを推定するには、以下の式でナイーブベイズ識別を行う。

$$C_i^* = \arg \max_{c_i} P(c_i|L) \quad (12)$$

提案手法では、(12) 式で得られた C_i^* が撮影された位置を示している。

4. 実験

まず、本実験における問題の設定を行う。位置情報付き参照画像が準備してあるとする。各参照用画像には、撮影位置 $l (l = 1, \dots, L)$ 、撮影方向 $d (d = 1, \dots, D)$ が記録されている。このとき、入力用画像 X について、撮影位置として適切な l と d を求めることを、本実験での問題とする。評価に用いるデータベースは、上記の問題設定に合致するデータベース2つを用いる。1つは独自に作成したデータベース1、もう1つは[4]で用いられているデータベース2である。これらのデータベースを用いることで、従来手法との比較実験を行う。以下でデータベースの詳細、実験概要、実験結果を示す。

4.1 評価用データベース

データベース1 参照用画像は大学キャンパス屋外にてGPS機能付きデジタルカメラにより、画像を撮影することにより作成した。問題設定に対応するように、 $L = 8$ 地点、 $D = 4$ 方向の画像を2010年2月に撮影し、各方向に対し撮影位置、撮影方向に位置ずれを生じさせて25枚の画像を撮影した。そのため計800枚となる。次に、入力用画像は参照用画像と同様の機器を用いて撮影時期、時間を変えて撮影を行った。入力用画像も参照用画像作成に用いた、地点 l 方向 d に属するように撮影した。入力用画像は11月に撮影されたもの32枚、1月に撮影されたもの20枚、2月に撮影されたもの298枚で構成されている。図7(a)にデータベースの一部を示す。これら

の画像にはGPS機能によりジオタグが付与されているが、評価値には用いず地図上へのマッピング実験にのみ用いる。

データベース2 小野らが用いたデータベースは、大学キャンパス屋外 $L = 19$ 地点、 $D = 60$ 方向分の画像を、同位置で異なる日時で撮影した2つのセットを構築している。それらの画像は、三脚で高さを固定したカメラでその地点を中心に一周をなるべく等間隔に分割するようにして撮影されたものである。そのため、単一方向には1つの画像しか存在せず、データベース1と趣向が異なる。そこで、実験では60方向分の画像を、15枚で1方向とすることで求める方向を4方向に制限した。つまり同定する対象は $L = 19$ 、 $D = 4$ となる。図7(b)にデータベースの一部を示す。

4.2 実験概要

従来法との比較実験を行う。従来法は、データベースの規模や前提条件が異なるが、[3]の文献で用いられた7つの特徴量を用いた方法、投票を用いたSIFTの2つを用いる。正誤判定は同定された l 、 d が完全に一致した場合を正解とする。ここで、実験パラメータについて述べる。利用する画像サイズを 640×425 とし、 r は最大を50に設定した。Randomized Treesの学習パラメータについては表1にまとめる。これらのパラメータを用いて、従来法との比較実験を行う。

表1 Randomized Treesの学習パラメータ

項目	パラメータ
決定木の数	25
決定木の深さ	7
特徴選択回数	400
しきい値選択回数	5
サブセットの割合	20%

まず、データベース1とデータベース2を用いて提案手法の汎用性を評価する。このとき、本実験ではランキングされた位置に対する正解率を評価する。次に、同定時間の比較を行う。ここで比較するのは同定にかかる時間であり、特徴抽出の時間は考慮しない。

次に、提案手法の特徴量算出領域サイズを決定するパラメータ r の同定精度への影響、提案手法で用いる4つの特徴量、式5~7の各自の同定精度への影響を調査する。そして、最後にデータベース1を用いた地図上へのマッピング実験を行う。

4.3 実験結果

4.3.1 提案手法の汎用性の評価

まず、データベース1における実験結果を図8に示す。図8は横軸に同定された順位に対する累積正解率、縦軸に正解率を取ったものでありグラフが左上に近くなるほど性能が高いことを示す。結果より、IM2GPS、提案手法、SIFTの順の性能が得られた。まず、SIFTの性能が最も低い要因としては、データベース1は植物を多く含



図 7 データベースの一部

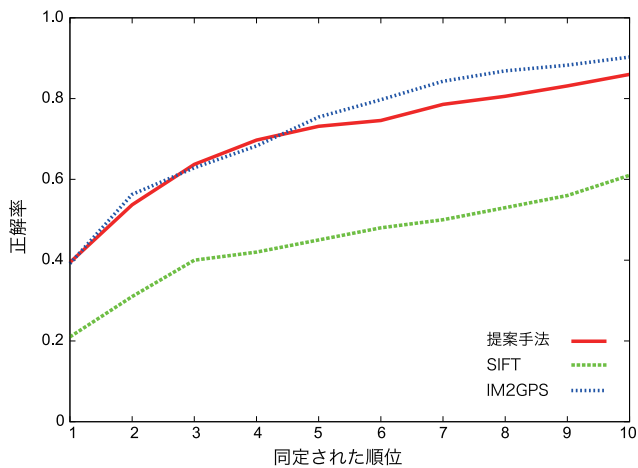


図 8 データベース 1 における累積正解率

む画像が多いということが挙げられる。植物から多くの特徴点が検出されることが悪影響し、精度低下を招いていると考えられる。次に、IM2GPS は提案手法より性能が良いという結果が得られた。これは、データベース 1 の入力画像が参照用画像が各方向ごとに位置ずれが大きいものも含むためであると考えられる。IM2GPS はシーン全体を記述する特徴量として、カラーヒストグラム、Texton ヒストグラム、Line Features などヒストグラムを用いた特徴量を 3 種類用いているため、やや位置ずれに強い記述を行うことができる。それに対し、提案手法は GC により得られる確信度画像の画像座標を、直接参照する特徴を用いており、領域から特徴量を算出することで少しの位置ずれには強いが、大幅な位置ずれに対しては対応が難しいと考えられる。

次に、データベース 2 における実験結果を図 9 に示す。データベース 2 は参照用画像と入力画像の位置ずれがほとんど無い。これに対し、撮影日時の違いにより大きな照明変動がデータベース 2 における問題である。結果より SIFT、提案手法、IM2GPS の順の性能が得られた。まず、SIFT が最も性能が良かった理由としては、データベース 1 とは違い植物が存在しないシーンも多くあるという点、SIFT の照明変動に対する頑健さが挙げられる。また、提案手法は IM2GPS より性能が良いという結果が得られた。これは、データベース 2 が参照用画像

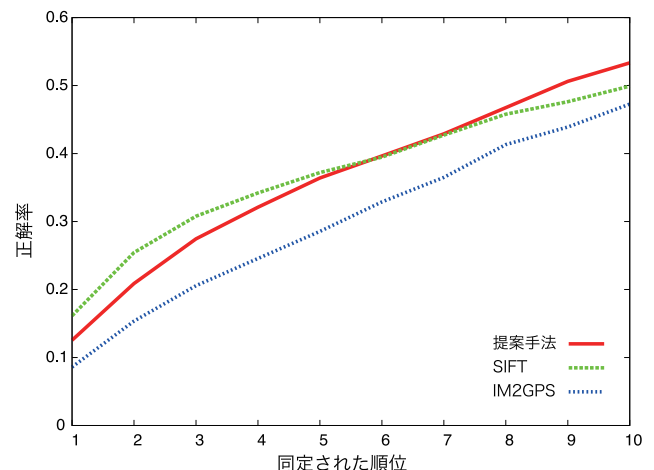


図 9 データベース 2 における累積正解率

と入力用画像の位置ずれが非常に少ないため、画像座標上の変化が少ないことが大きな要因であると考えられる。他、提案手法で用いている GC による照明変動の吸収も考えられる。これに対し、IM2GPS はシーン全体を記述するため大きな照明変動がシーン全体に起こることで、精度に影響を与えたと考えられる。

4.3.2 領域サイズの影響

提案手法の GC 特徴算出に用いる領域サイズを決定するパラメータ r の、同定精度への影響を調査する。調査には、データベース 1 を用いる。調査項目は r を 10~50 を固定して用いたもの、また最大値を 50 にし、学習により選択させたものの計 5 項目の事例を調査する。図 10 に調査結果を示す。結果より r を大きくしていくと同定精度が下がっていくことが分かる。これは、 r を大きくするほどシーンの詳細を記述することが困難となるためであると考えられる。また、 r に最大値を設け学習により選択させることで固定の r を用いるよりも性能が向上することが分かる。これは、大域的に捉えるべき領域と局所的に捉えるべき領域が学習により自動決定され、シーンを頑健に記述できるためであると考えられる。

4.3.3 特徴量に対する同定精度の比較

提案手法で用いている特徴量、式 5~7 の同定精度への影響を調査する。調査にはデータベース 1 を用いる。

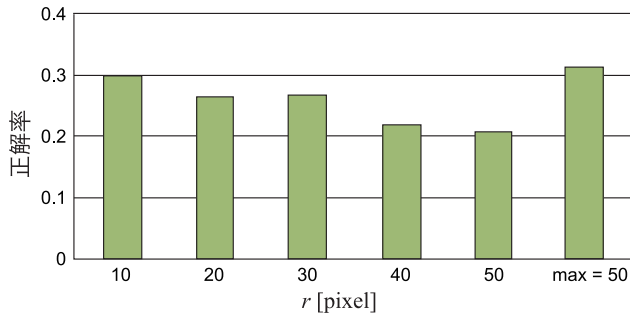


図 10 パラメータ r に対する同定精度

図 11 に、各特徴量のみを用いた場合の同定精度と、すべての特徴量を用いた場合の同定精度を示す。結果より、

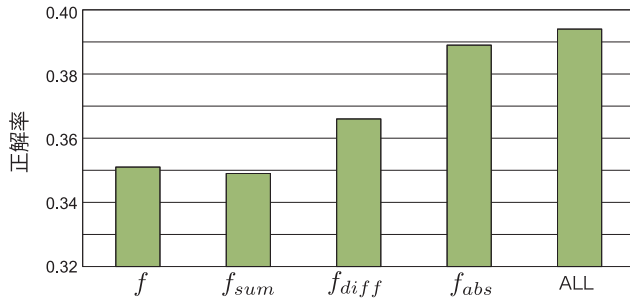


図 11 特徴量に対する同定精度の比較

ある領域のみを用いる f よりも領域同士の組み合わせを利用した f_{diff} や f_{abs} の方が同定精度が高くなっていることがわかる。これにより、共起を用いることでシーンを頑健に記述できると考えられる。また、それらを単体で用いるより 4 つすべてを用いた場合の方が、同定精度が高くなることが分かり、多くのシーンを記述できていると考えられる。

4.3.4 同定時間の比較

各手法における同定時間を表 2 に示す。これから、SIFT を用いた方法が最も時間がかかることが分かり、提案手法が最も時間がかからないということが分かる。さらに、SIFT より高速である IM2GPS と提案手法を比較すると、約 10^3 倍高速であることが分かる。これは、従来法のように全特徴次元を用いた全探索を行うのではなく、Randomized Trees の木構造を用いたノードの探索により、特徴空間を階層的に探索できるためであると考えられる。

表 2 各種法における同定時間

手法	同定時間 [ms]
提案手法	0.6
IM2GPS	6.3×10^2
SIFT	7.5×10^3

4.3.5 地図上へのマッピング

提案手法により同定された位置における参照用画像の中から代表画像を選び、地図上へのマッピングを行う。マッピング結果を図 13 に示す。 $l=6, d=1$ の地点においては、積雪による影響を受けず撮影位置を同定できていることが分かる。また、 $l=5, d=2$ の地点の入力画像には画像の左部分の植物の葉が存在している。これに

対し、参照用画像は季節がことなり左部分に存在していた葉が枯れてしまっているが、これを捉えることができ撮影位置を同定できている。最後に、 $l=7, d=1$ の地点では 2 位に正解位置がある。これは、1 位に選ばれた画像の構造と入力画像の構造が非常に類似しているため誤同定が起こったと考えられる。

4.4 考 察

Randomized Trees により選択された特徴量について考察を行う。まず、図 12 に木構造の各深さごとに選択された GC の種類とその割合を示す。図 12 より、木構造の上層ノード (第 1 層) では、植物が多く選択されている。一方、地面は植物とは異なり最下層ノード (第 11 層) で多く選択されている。最下層ノードでは空や植物は選択されておらず、垂直物と地面を利用して分類のためのシーンの詳細を記述していることが分かる。

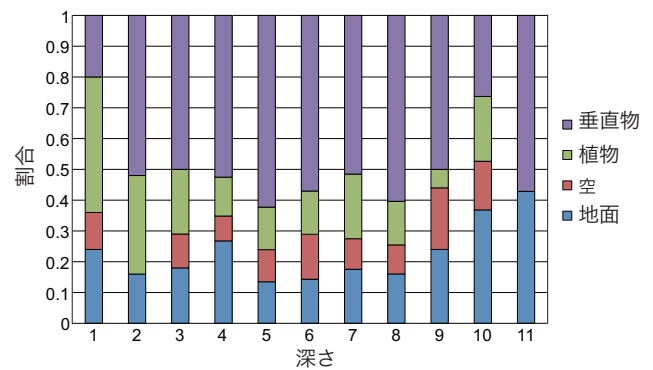


図 12 階層ごとの選択された確信度

5. おわりに

本稿では、画像から Geometric Context の確信度を用いて共起特徴を算出することによりシーンを記述し、それを撮影位置が関連づけられた Randomized Trees により識別することにより、画像の撮影位置を同定する手法を提案した。評価実験により、従来手法と比べ約 10^3 倍高速であることを示した。

提案手法の問題点としては、Geometric Context を用いていることにより、対応可能なシーンが野外に限定される点、また、Randomized Trees の学習がデータベースの増大に従って困難になる点が上げられる。しかし、対象が野外シーンに限定される問題では、シーンを記述することが可能な特徴量である Gist [19] などを用いることにより、対象の範囲を拡大することが可能である。また、データベースの増大に対しては、Randomized Trees をオンラインで学習する手法が提案されており [15]、これを用いることでの対応が期待される。

今後は、[3] で用いられているデータベースと同等のデータベースでの評価実験を行う予定である。

謝 辞

本論文の第 4. 章の実験では九州大学大学院システム情報科学府小野善太郎氏、九州大学大学院システム情報科学研究院内田誠一先生にデータベースを提供して頂いた。ここに感謝の意を表する。



図 13 提案手法による同定結果

文 献

- [1] L. Breiman, "Random forests", Machine learning, Springer, 2001, 45, 5-32
- [2] J. Shotton, M. Johnson and R. Cipolla. "Semantic Texton Forests for Image Categorization and Segmentation". In Proc. CVPR, pp. 1.8, 2008.
- [3] James Hays and Alexei A. Efros, "IM2GPS:estimating geographic information from a single image", In Proc. CVPR, pp.1-8, 2008.
- [4] 小野善太郎, 堀田政二, 馮堯楷, 内田誠一, "局所特徴量を用いた事例参照に基づく位置同定," 信学技報, 2009,212,pp31-36
- [5] 上野智史, 橋本真幸, 米山暁夫, 川田亮一, "SIFTを用いた都市環境の建物画像検索によるユーザ位置・方向推定手法に関する一検討" 信学技報, 2009,109(306),pp.127-132
- [6] 近藤和樹, 堀田政二, "ブロック置き換えとブロック削減を利用したカラー画像認識," 第 14 回画像センシングシンポジウム (SSII09) 論文集, IN1-10,Jun,2008
- [7] Lepetit, V., Fua, P. "Keypoint Recognition Using Randomized Trees" IEEE Trans. Pattern Anal. Mach. Intell., IEEE Computer Society, 2006, 28, 1465-1479
- [8] Csurka, G., Dance, C.R., Fan, L., Willamowski, J. and Bray, C., "Visual categorization with bags of keypoints", ECCV International Workshop on Statistical Learning in Computer Vision (2004).
- [9] YAHOO. Flickr. <http://www.Flickr.com/>.
- [10] D. Lowe, "Distinctive image features from scaleinvariant keypoints", Int. Journal of Computer Vision, 60(2), pp. 91-110, 2004.
- [11] D. Hoiem, A. A. Efros and M. Hebert, "Geometric context from a single image", In Proc. ICCV, 1, pp. 654-661 (2005).
- [12] 薄充孝, 中川知香, 佐藤智和, 横矢直和 "ランドマークデータベースに基づく静止画像からのカメラ位置・姿勢推定," 日本バーチャルリアリティ学会論文誌, Vol.13 No.2 2008.
- [13] R. Cipolla, D. Robertson and B. Tordoff, "Image-Based Localization," In Proc. VSMM, pp. 22-29, 2004.
- [14] R.Szeliski. "Where am I?":ICCV2005 Computer Vision Contest. <http://research.microsoft.com/iccv2005/contest/>.
- [15] Osman Hassab Elgawi, "Online random forests based on CorrFS and CorrBE," CVPR Workshops, 2008, pp.1-7, 2008
- [16] W. Zhang, J. Kosecka, "Image Based Localization in Urban Environments," In Proc. 3DPVT,33-40, 2006
- [17] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features", In Proc.ECCV.,2006
- [18] P.Felzenswalb and D. Huttenlocher: "Efficient Graph-based Image Segmentation". Int. Journal of Computer Vision. Vol.59 No2, pp. 167-181 (2004).
- [19] A. Oliva and A. Torralba. "Building the gist of a scean: The role of global image features in recognition", In *Visual Perception, Progress in Brain Research*, Vol.155, 2006.