

画像記述子の線形結合による ヴィジュアルローカリゼーション

鳥居 秋彦^{†1} Josef Sivic^{†2}
Tomas Pajdla^{†3} 奥 富 正 敏^{†1}

GPS などから得られた位置情報を持つ大規模な画像データベースが与えられた場合に、クエリ画像の位置情報を推定するアルゴリズムを提案する。本研究の貢献は以下の3点にある：(1) イメージグラフ上での回帰問題としての画像ベースローカリゼーションの定式化；(2) データベース画像ペア（グラフのエッジ）に対する画像検索アルゴリズムの設計；(3) 検索されたデータベース画像ペアの位置情報補間によるクエリ画像位置推定。これらにより、計算コストの高い特徴点対応付け・幾何モデル推定を行うことなく、高速な位置推定が可能である。アルゴリズムの性能評価として、Google Street View Pittsburgh データセットを用いた実験を行う。

1. はじめに

本研究のゴールは、GPS などから得られた位置情報（ジオタグ）を持つ大規模な画像データベースを用いて、クエリ画像の位置情報を推定する、ヴィジュアルローカリゼーション（以下ローカリゼーションと呼ぶ）である。クエリ画像とデータベース画像は、通常異なる日に撮影されるため、視点、照明変化、遮蔽など様々な問題を含む挑戦的な課題である。

ジオタグ画像データベースに対して、クエリ画像のある特徴から地図上位置へのマッピングを導く、回帰的アプローチによる位置推定手法を考える。どのような回帰の形式をとるかの選択は重要であり、画像特徴表現、類似度計測、地図の構造、データベースサイズなど、様々な要因に依存する。

本研究ではヴィジュアルローカリゼーションを2段階の回帰問題として定式化する。第1ステップでは、データベース画像からクエリと同じ場所を示す（つまり同じ3Dシーンを

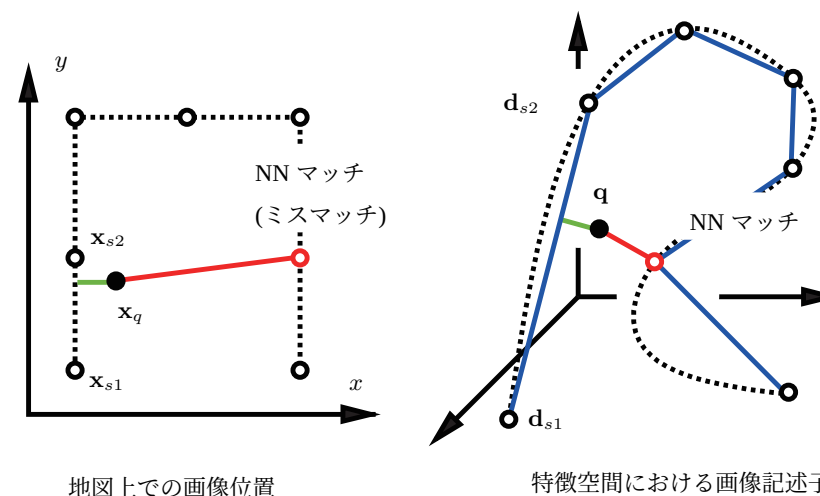


図1 イメージグラフから得られる隣接画像の線形結合マッチング。左図：2D 地図上のパスに沿ってローカライズされた画像。データベース画像はエッジでつながれ（破線）イメージグラフを定義する。黒いマーカーはクエリ画像位置を表す。右図：左図に対応する特徴空間での画像記述子。グラフのエッジに沿って構成される画像ペアによる1D部分空間（青線）とクエリ記述子との距離で類似度を測ることで、マッチング精度の向上が期待できる。赤線は画像記述子のみを用い、独立にマッチング（最近傍探索）を行った場合の、ローカリゼーション失敗例を表す。

撮影した) 部分集合画像: *surrounding image set* (包囲画像集合) を探索する。第2ステップでは、対応した包囲画像集合のGPS情報の補間によりクエリ画像位置を推定する。

第1ステップでは、大規模画像/物体認識^{5),14),22),23)} や位置認識^{6),15),28)} において顕著な性能を示している bag of features (BoF) を用いる。さらに、画像をノード、地図上で隣接する画像をエッジで接続した *image graph* (イメージグラフ)^{17),25),33),37)} としてジオタグ画像データベースを表現する。このイメージグラフからデータベース画像ペアを抽出し、画像ペア BoF ベクトルの線形結合を考慮した類似度計算を行う。イメージグラフ上の複数特徴ベクトルの線形結合を考慮することで、欠損またはミスマッチした局所画像特徴量に対するロバスト性を向上する。視点変化と照明変化に対しより適応的になることから、マッチングの精度の向上が期待できる。一方、クエリ時の計算コストの増分はごく僅かである。

クエリ画像と同じコンテンツを含む画像集合が与えられれば、クエリの位置姿勢は *structure*

^{†1} 東京工業大学

^{†2} INRIA, WILLOW project, Laboratoire d'Informatique de l'École Normale Supérieure, ENS/INRIA/CNRS UMR 8548.

^{†3} CMP, CTU in Prague

from motion(SfM)^{8),32)}を用いて推定できる。SfM手法は成熟しつつあるが基本的に計算量コストが高く、良い初期値を必要とする。そこで、提案するローカリゼーションの第2ステップとして、クエリ画像とマッチしたデータベース画像ペア間で相対的類似度を計算し、GPS位置補間を行うことで、クエリ位置を推定する。

本研究の貢献は以下の3点にある：(1) イメージグラフ上での回帰問題としての画像ベースのローカリゼーション定式化；(2) データベース画像ペア（グラフのエッジ）に対する画像検索アルゴリズムの設計；(3) 検索されたデータベース画像ペアの位置補間によるクエリ画像位置推定。

関連研究: 画像検索アルゴリズム^{6),15),26),28)}と、幾何学的検証 (geometric verification) アルゴリズム^{1),11),18)}に基づいた、大規模ローカリゼーション研究は目覚ましい進展をとげている。

Cummins-Newman⁶⁾, Knopp ら¹⁵⁾は BoF を用いた大規模アピランススペースのローカリゼーションを提案した。これらは、各データベース画像に対するマッチングを行うもので、BoF ベクトルの線形結合を考慮したものではない。Schindler ら²⁸⁾は密な画像列から vocabulary tree²²⁾を用い、クエリ画像の特徴を地図上の隣接視点にまたがった投票を行うことで、画像ベースのローカリゼーションを実現した。Zamir-Shah³⁵⁾は、クエリ画像の局所特徴量と Google Street View 画像データベースの局所特徴量の (visual word への量子化を介さず) マッチングを行い、2D 地図上で対応点の分布を解析する形でローカリゼーションを行った。

本研究で提案する画像類似度では、特徴記述子の線形結合を考慮するため、特定物体認識において非常に有効な手法であるクエリ拡張^{4),5)}と似た性質を持つ。我々は、全ての線形結合を考慮しつつ、クエリでなくデータベースを拡張する。そして検索時には閉型式での高速計算が可能であるため、付加的な BoF ベクトルや記述子の計算⁵⁾やデータベース側における付加データ保存の必要がない。

Irschara らは、SfM から得られる3D点群を用いた画像記述子合成により、撮影されていない視点を拡張することで、高精度なローカリゼーションを行った¹¹⁾。Li らは多くのデータベース画像間で頻繁現れ、さらに、幾何学的検証を通して正しく対応付けられた3D特徴点を用いた、優先的マッチングに基づいたローカリゼーション手法を提案した¹⁸⁾。本研究では、計算コストの高い3D復元や、常に得られるとは限らない3D特徴情報を用いず、データベース画像間の位置補間によるローカリゼーションを目的とする。さらに、ビデオシーケンス¹¹⁾やランドマーク画像集合¹⁸⁾ではなく、Google Street View に代表される一般的な街路画像列に焦点を当てる。

特徴ベクトルの組み合わせを考慮した画像類似度計算によって、データベースをより疎に構成可能であれば、データベースサイズの縮小につながる。従って、我々のアプローチは、検索とローカリゼーションのための画像データセットの圧縮手法に関連がある。関連研究として、イメージグラフ^{17),25),33),37)}、画像要約²¹⁾、3D点表現^{11),18)}、識別的な特徴^{15),16),28)}などを用いた手法が挙げられる。

我々の回帰手法が標準的なアプローチ³⁾と異なる点は、特徴空間でのデータ点類似度のみを取り扱うのではなく、地図上で画像が整理されている利点を生かし、イメージグラフから得られる空間的に隣接する画像ペアの線形結合を考慮する、そして、その構造を直接回帰問題に埋め込んだ所にある。

2. イメージグラフを用いた回帰問題としてのローカリゼーション

回帰問題としてのローカリゼーションタスクを定式化し、提案する回帰関数の2ステップアプローチについて概略する。

本研究では、画像を tf-idf 重み付き²⁷⁾BoF 記述子 $\mathbf{d}$ として表現する²⁷⁾。Visual vocabulary のサイズによって、記述子ベクトルの次元は決まり、ここでは 10^4 から 10^6 程度のものを用いる。

グラフ $G = (D, E)$ として構成される**イメージグラフ**を考える。 N 個のノード $D = \{(\mathbf{d}_i, \mathbf{x}_i)\}_{i=1}^N$ は画像記述子 $\mathbf{d}$ と対応する地図上2D位置 $\mathbf{x}_i$ によって構成される。グラフ G のエッジ E は、同じ対象を共有する隣接視点を接続する。例えば、Google Street View データでは、 D はパノラマ画像（の記述子と位置）、 E はある画像から隣の画像へとナビゲートさせるリンクに相当する。

ここで、イメージグラフ G を用いて、記述子 $\mathbf{q}$ の位置 $\mathbf{x}_q$ を与えるローカリゼーション関数 f を考える

$$\mathbf{x}_q = f(\mathbf{q}, G). \quad (1)$$

エッジ E は画像ベースのローカリゼーションにおいて有用な構造を与え、個々の画像記述子を用いるよりも、高精度なローカリゼーションが可能であることを示す。エッジによってリンクされている記述子ペア $\mathbf{d}_1, \mathbf{d}_2$ （すなわちイメージグラフ内で隣接する視点）を考える。ここで、クエリ画像が地図上のデータベース画像ペアの中間位置で撮影されたとする。その場合、クエリ画像の記述子 $\mathbf{q}$ は、両方の記述子とコンテンツを共有するであろう。個々に類似度を測るよりも、 $\mathbf{d}_1, \mathbf{d}_2$ を結合した記述子の方が $\mathbf{q}$ と高い類似度を持つことが容易に想像できる。本研究では $\mathbf{d}_1, \mathbf{d}_2$ の線形結合が有意義な表現であり、ローカリゼー

ションの性能を向上することを証明する。イメージグラフ G とクエリ集合 D_Q が与えられた場合における、適切なローカリゼーション関数 f の設計に焦点を当てる。5章の実験結果では、イメージグラフの密度（データベース密度）変化に対するローカリゼーション性能評価を行う。

イメージグラフを用いた2ステップ回帰手法: クエリ画像とビジュアルコンテンツを共有する**包囲画像**のみを用いた、クエリ画像の**相対的な**位置推定を目的とする。利用する包囲画像の数は、環境の3D構造に依存するが、典型的な都市環境ではそれほど多くない。

この枠組において、 f を2ステップ構成とするのは自然である。第1ステップでは、クエリとコンテンツを共有する包囲画像(の部分集合)を識別し、そして第2ステップでは、包囲画像に対して相対的なクエリ位置を計算する。特に画像認識において、画像検索手法^{(12),(22),(23)}と幾何学的判定⁽²³⁾という確立された2ステップアプローチが存在することからも、本研究で提案する2ステップアプローチは有効であると考えられる。

ここに**線形回帰**ローカリゼーション関数 f

$$\mathbf{x}_q = \sum_{s \in S(\mathbf{q})} w_s \mathbf{x}_s \quad (2)$$

を定義する。ここで $\mathbf{x}_q$ は予測されたクエリ位置であり、

$$S(\mathbf{q}) = f_S(\mathbf{q}, G) \quad (3)$$

はイメージグラフ G の部分集合であり、**surrounding set retrieval (包囲集合検索)** 関数 f_S によって決定される **surrounding sets(包囲集合)** である。

次節ではクエリベクトル $\mathbf{q}$ に対する包囲集合の決定法、4章では、特定の包囲集合に対する予測推定重み w_s の計算法について述べる。

3. 画像ペア線形結合を考慮した類似度計算

本節では、イメージグラフが与えられた場合の包囲画像集合を定義し、エッジ E によってリンクされた画像の線形結合を考慮した類似度計算による検索法の詳細を述べる。

包囲画像集合. クエリ $\mathbf{q}$ が与えられた場合、 $\mathbf{q}$ とビジュアルコンテンツを共有する画像 D_S を内包し、イメージグラフ G の部分集合グラフ $S(\mathbf{q}) = (D_S, E_S)$ を決定することが、我々の課題である。

クエリ画像記述子ベクトル $\mathbf{q}$ は、 $S(\mathbf{q})$ 内のエッジによってリンクされた画像記述子ベクトルの線形結合として、近似できることを経験的に導いた。このアプローチの背景には2つの要素がある。ひとつは、クエリベクトルには存在する **visual word** が、遮蔽や検出誤差、

量子化誤差などの影響で、部分集合内のあるデータベース画像では欠損している場合、複数画像を考慮することで“補充”することができる。もうひとつは、特徴ベクトルの線形結合は **view interpolation (視点補間)** として捉えることができる。しかし、ここでは高次元特徴空間での補間であり、データベースにはない中間視点として効果的に拡張することになる。

上述したモチベーションのもと、 G 内のエッジひとつで構成された包囲集合 $S(\mathbf{q})$ の探索を行う。イメージグラフの例として、グラフ上の距離⁷⁾が短い画像がコンテンツを共有するように平面グラフが用意された **Google Street View** 画像データベースが挙げられる。 $S(\mathbf{q})$ に**単位グラフ距離の画像ペア**という制限を与えることで、 $S(\mathbf{q})$ に含まれる画像がビジュアルコンテンツを共有するという仮定が十分に成立し、線形回帰子(2)の組み合わせを意義のあるものにする。街路に沿って画像列が撮影されており、空間的に近い画像がコンテンツを共有するデータからのローカリゼーションにおいて、この選択は特に適切である。

最近傍探索 (nearest neighbor: NN) によるマッチングは地図データベースが密な場合には実用的である。しかしながら、(データベース画像間の各距離が長い) 疎なデータベースにおいては、短いグラフ距離で表現される画像ペアを検索の方が効果的であることを、5節の実験で明らかにする。これらの考察の直感的な解釈を図1に記す。

ここで、クエリ $\mathbf{q}$ に最も整合するペア $S(\mathbf{q})$ は、 E で定義される全てのペアの中で最も類似するアフィン結合を探索することで得られる、

$$S(\mathbf{q}) = \operatorname{argmin}_{(i,j) \in E} \min_{\alpha \in \mathbb{R}} \|\mathbf{q} - ((1-\alpha)\mathbf{d}_i + \alpha\mathbf{d}_j)\|^2, \quad (4)$$

α はアフィン結合を表すパラメータである。 α に関する最小値は以下の閉型式として得られる、

$$\alpha^* = \frac{(\mathbf{d}_j - \mathbf{d}_i)^\top (\mathbf{q} - \mathbf{d}_i)}{\|\mathbf{d}_j - \mathbf{d}_i\|^2}. \quad (5)$$

線形結合と凸結合も考えられるが、経験的には、アフィンと凸結合との差は殆ど無く、アフィンと線形間の性能差も1%程度であった。アフィン(と凸)結合の利点、つまり結合をひとつのパラメータで表現できることが、つづく位置推定ステップ(4節)の1次元線形回帰を容易にする。

式(4)を最小値にする α が閉型式で得られることから、特徴ベクトルの線形結合を考慮した探索でも、計算コストの増加はごく僅かである。同時に、転置インデックス⁽³⁴⁾を用いた検索^{(23),(31),(34)}手法も応用可能である。さらに、典型的なイメージグラフにおいて考えられる画像ペアの数(イメージグラフのエッジ数)は限られておりイメージグラフの画像数に対して高々線形である。

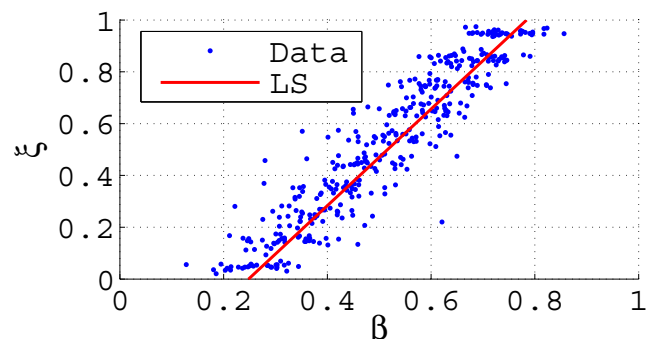


図2 相対類似度 β と位置補間パラメータ ξ (式 (6)). イメージグラフの画像ペア標本から学習した, β と位置補間パラメータ ξ のアフィン従属性を確認できる.

次節では, クエリに対する包囲集合 $S(\mathbf{q})$ が得られた次のステップであるクエリ位置推定について述べる.

4. クエリ画像位置推定

本節では, 式 (2) における位置推定の重み w_s 算出法について述べる. α^* を位置予測推定にそのまま使うことが自然であるが, 実際には, α^* の値が中央 (0.5 付近) に偏りがちである. 原因として, 遮蔽による検出ミスや局所特徴記述子の不変性を越えた変化などが考えられる. クエリ画像がデータベース画像と全く同じ位置, しかしながら異なる時刻に撮影された場合, これらの影響によって $\mathbf{q} \neq \mathbf{d}_1$ つまり $\alpha \neq 0$ となってしまう (理想的には $w_1 = 1$ かつ $w_2 = 0$).

より良い位置推定は下記の方法で得られることを経験的に導いた. はじめに, クエリ画像と最も類似したデータベース画像ペア間で, 相対的類似度を計算する,

$$\beta = \frac{\mathbf{q}^\top \mathbf{d}_{s1}}{\mathbf{q}^\top \mathbf{d}_{s1} + \mathbf{q}^\top \mathbf{d}_{s2}}. \quad (6)$$

各クエリ $\mathbf{q}$ とそれに対応する包囲集合 $S(\mathbf{q}) = (s_1, s_2)$ を選び, $\mathbf{q}$ の位置を包囲集合のアフィン結合で表現することで, 位置補間パラメータ ξ が得られる,

$$\mathbf{x}_q = \mathbf{x}_{s1} + \xi (\mathbf{x}_{s2} - \mathbf{x}_{s1}). \quad (7)$$

真値として $\mathbf{x}_{s1}$, $\mathbf{x}_{s2}$, $\mathbf{x}_q$ が与えられた場合, 式 (6) の画像類似度から得られる β に対する ξ の値を図2に示す. β に対する ξ のアフィン結合性を明確に確認できることから, アフィ

ン補間によって β から ξ を推定する.

$$\xi_q = a_0 + a_1 \beta. \quad (8)$$

これらの考察を基に, 独立な学習用データセットから a_0, a_1 を計算し, $\mathbf{x}_q$ の位置を以下の式によって推定する.

$$\mathbf{x} = \mathbf{x}_{s1} + (a_0 + a_1 \beta)(\mathbf{x}_{s2} - \mathbf{x}_{s1}) \text{ with } (s_1, s_2) = S(\mathbf{q}) \quad (9)$$

5. 実験

イメージグラフのエッジから得られる画像記述子ペアを考慮した, マッチングと位置推定の性能評価を行う.

マッチング性能評価. はじめに, Google Street View Pittsburgh Research データセット ("Street View Research") を用いた実験を行う. このデータは 8,999 枚のパノラマ画像で構成される. 検出される特徴点の数を調整するために, 各パノラマ画像は $1,664 \times 512$ 画素にダウンサンプルした (オリジナルサイズの約半分).

画像は BoF によって記述する^{23),31)}:

- (1) SURF 特徴量の検出と記述²⁾,
- (2) 450 枚の画像をランダムに選択, approximate k-means clustering^{20),23)} を用い, visual vocabulary を構築,
- (3) visual vocabulary を用いた SURF 記述子の量子化, データベース中の語彙頻度に基づいた重み付けによる tf-idf ベクトル算出²⁷⁾,
- (4) tf-idf ベクトルの (単位距離) 正規化,
- (5) tf-idf ベクトルの内積として画像間類似度の算出

奇数番号フレーム画像を位置情報付きの画像データベース D , 偶数番号フレーム画像をクエリセット D_Q として用意する. グラフ G で表されるイメージグラフは, データベース画像 D から各 Δ 番目の画像を標本し構築する. この Δ をサンプリングレートと呼ぶ. サンプリングされたデータベース画像内で, フレーム番号の連続するものをグラフのエッジで結ぶ. Street View Research データセットでは各画像間距離は約 1 メートルであるため, サンプリングレート $\in \{5, 10, 15, 20, 25\}$ は, それぞれ $\{10, 20, 30, 40, 50\}$ メートル間隔でデータベース画像が配置されていることに相当する (データベース画像セットは奇数番号のみのため, オリジナルデータの半分であることを注意しておく).

以上のデータセット構築方法により, クエリ画像セット D_Q は, グラフ G 内のひとつのエッジに沿っており, クエリに対する真値となる包囲集合を定義する. マッチング性能評価

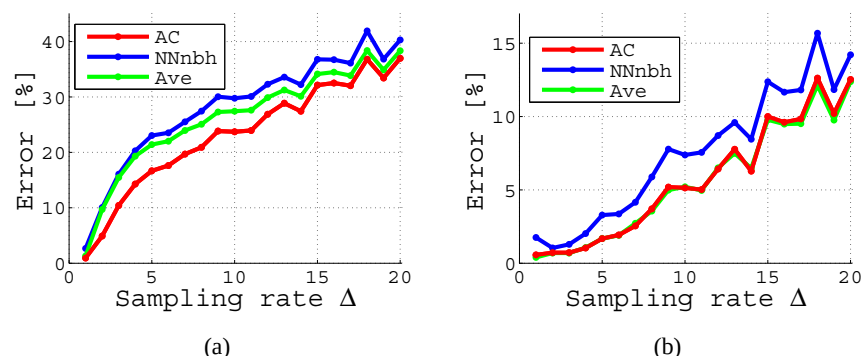


図3 マッチングの実験評価。赤 (AC), 緑 (Ave), 青 (NNnbh)。エラーは、データベース画像の異なるサンプリングレート Δ (x 軸) においてミスマッチしたクエリ画像の割合 (y 軸) で定義した。ミスマッチは以下の 2 つの基準で評価した: (a) クエリが真値のペアに正しく割り当てられた, (b) 検索結果エッジのノード (ペア) のひとつが少なくとも真値のノードペアと一致する ((a) の完全一致よりゆるい条件である。) (a) において, AC による性能向上は明らかである。(AC) と (Ave) は粗いレベルでのマッチングにおいて, 同様の結果となったが, どちらもベースラインの (NNnbh) よりも性能が良いことに注意されたい。

のゴールは, 各クエリに対して正しい包囲集合を探索・検出することである。線形結合によるマッチング (AC), 画像ペア類似度の平均によるマッチング (Ave), ベースラインとなる最近傍探索 (NN) との間で実験評価する。ここで, 標準的な NN の検索結果では, 画像ペアを出力しない。画像ペア検索結果として公平評価をするために, NN で最も類似したデータベース画像を含む包囲集合内で, 2 番目にクエリ画像と類似するものを選び, 画像ペア検索結果として出力することにした (NNnbh)。サンプリングレート Δ を変化させて, データベースセットの密度を変え実験を行った結果を図 3 に示す。

次の実験では, 1,143 枚の Pittsburgh Google Street View 画像 ("Street View Web") をデータベースとして用いた。イメージグラフ構造は, maps.google.com でナビゲーションする際に用いられる位置間リンク情報を用いて構築した。画像とリンク情報はどちらも maps.google.com からダウンロードした。Street View Web の各画像間の位置間隔は約 10~20 メートルであり, Street View Research におけるサンプリングレート 5~10 に相当する。クエリには, Street View Research データセット 8,999 枚全ての画像を用いた。9 割以上のクエリ画像が, 異なる時刻, 季節, 位置で撮影されているため, この実験は現実的かつ挑戦的な問題設定である。最近傍探索 (NN), その拡張 (NNnbh), 平均を用いた探索 (Ave), 提案手法 (AC) との比較を図 4 に示す。

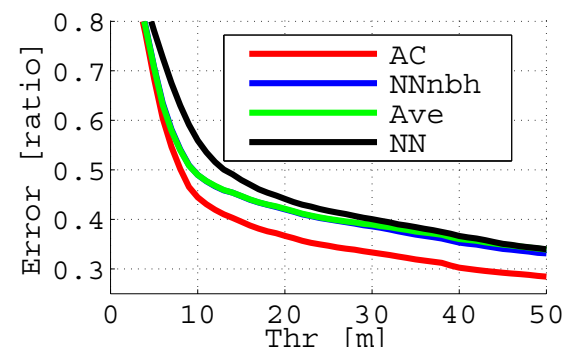


図4 Street view 画像ペアマッチング結果。マッチングによって割り当てたデータベース画像 GPS とクエリ画像 GPS (真値) の距離が閾値 "Thr" メートル (x 軸) 以上の場合, ミスマッチであると定めた。y 軸はマッチングに失敗したクエリ画像 ("Street View Research") の割合を表す。提案手法 (AC) による性能向上は明らかである。

位置推定. 続いて, Street View Research データセットを用いた, ローカリゼーションの精度評価を行う。これにより提案する 2 ステップ回帰によるローカリゼーションが完結する。

データは, 先ほどと同様データベース D とクエリセット D_Q に分ける。ここで, 最終 900 フレームを, 式 (8) 回帰パラメータの学習に用いた。ローカリゼーションの性能は, クエリ画像が真値とする GPS から距離 gt_{dist} 以内でローカライズされたかで判定する。性能は, このローカリゼーション精度の閾値を変化させつつ評価する。

(Ave) や学習無しなど, 他手法とも比較を行ったが, 結果的に (NN) が良い性能を示したため, 提案手法 (AC) とベースラインとなる (NN) 間で比較を示す (図 5)。 (NN) においては, クエリ位置はマッチしたデータベース画像の位置をそのまま割り当てた。

図 5 は, 提案手法 (AC) (y 軸) とベースライン手法 (NN) (x 軸) の性能を示す。曲線が破線のしたにある場合, (AC) の性能がベースラインの (NN) を上回ったことを意味する。サンプリングレート $\Delta = 5$ (青) と 10 (黒) の結果は, おおまかに, web 上の Google Street View 画像の間隔に対応する。小さな閾値 gt_{dist} , つまり, 誤差の許容範囲が小さい場合, AC による性能の向上が著しい。図 6 は, 第 2 ステップの回帰パラメータ学習効果を示す。図 7 は (AC) によって正しくローカリゼーションされたクエリ画像の例である。

ベースラインの NN に比べ, 提案手法が良い結果を示した理由は以下の 2 点にある: (1) 線形結合を考慮したマッチングによる検索性能の向上。 (2) 学習による位置補間推定による高精度な位置推定。

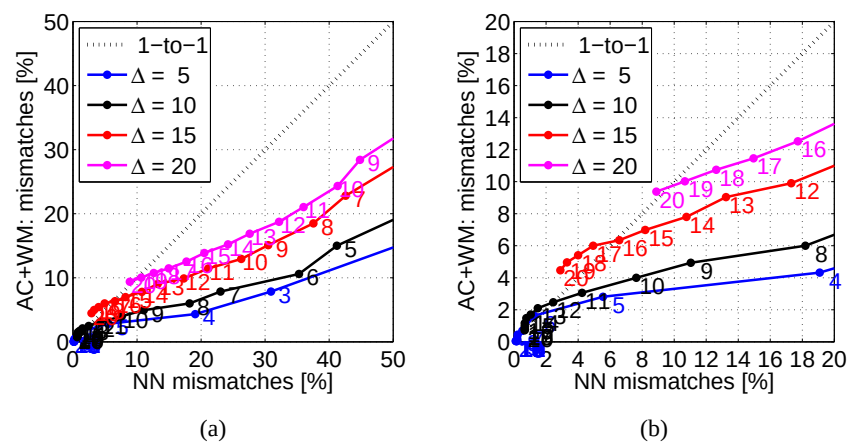


図 5 提案手法 (AC)(y 軸) とベースライン手法 (NN)(x 軸), それぞれのミスマッチの比率で表示したローカリゼーションエラー。それぞれの曲線は, ある固定したサンプリングレート Δ に対する精度を表す。曲線上のマーカとそれに対応する数値は, ローカリゼーション精度の許容範囲 (gt_{dist} メートル) を表す。(b) は (a) の拡大表示である。

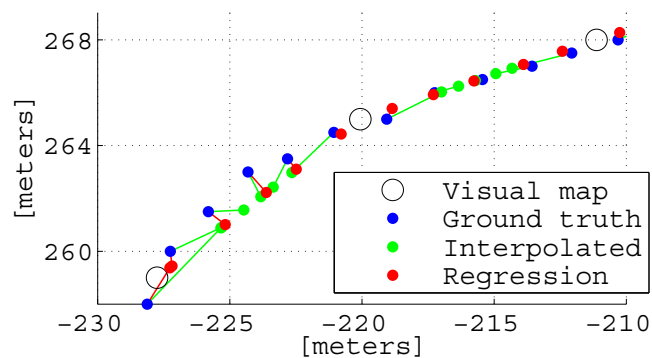


図 6 学習した回帰パラメータによるローカリゼーション精度向上例。黒“o”はデータベース画像の GPS 位置を表す。各マーカは, クエリ画像の GPS 位置 (真値)(青), 相対類似度 β を用いて推定したクエリ位置 (緑), 学習した重みを用いて推定したクエリ位置 (赤), を表す。 β による直接推定 (緑) に比べ, 赤ドットの位置が真値 (青) に近づいたことから, 学習の効果を確認できる。

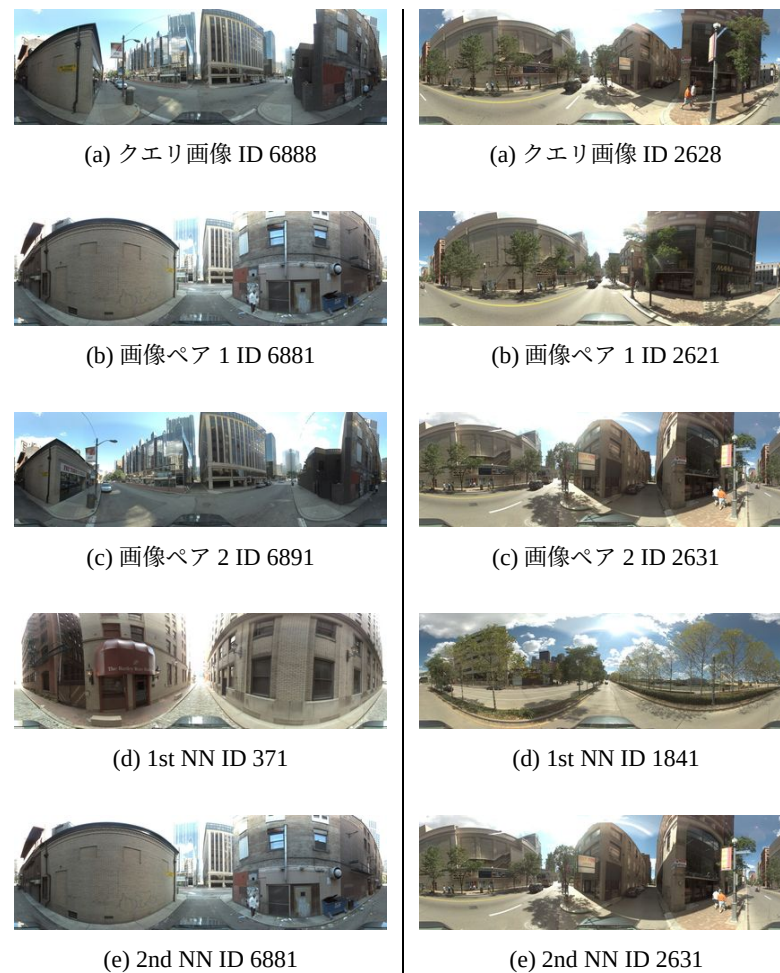


図 7 サンプリングレート $\Delta = 5$ でのローカリゼーション結果例。(a) クエリ画像。(b-c) 提案手法により選択された (正しい) 画像ペア。(d-e) 第 1, 第 2 最近傍探索結果の画像。(b) と (c) の間隔は約 10 メートルである。提案手法によって, クエリ画像位置は真値から 3 メートル以内に推定された。

スケーラビリティ. 画像ベースのローカリゼーションを実用するには、数万枚の画像を扱う必要がある。最先端の画像検索アプローチ^{12),22),23)}では、各画像記述子に対し独立に類似度を計算する。

本研究では、独立(クエリ画像 vslet@tokeneonedot データベース画像 1 枚)ごとに類似度を計算するのではなく、クエリ画像 vslet@tokeneonedot データベース画像ペアに対する類似度計算を行う。画像ペアをとることで検索すべき対象は増加するが、地図上の画像データという構造を考えるとその増加量はわずかである。イメージグラフ上では、各画像の持つ包囲集合の数が少なければ、ペアの数はノード数に対して線形である(傾きは13であろう)。例えば、データベース N 枚の画像が画像列として並べられていれば、考慮すべき画像ペア数は $N - 1$ である。完全に平面でのローカリゼーション問題では、画像ペアを3つ組み(triplet)に置き換えても、考慮すべき triplet の数は、(平面グラフの Euler 数より) 高々 $3N - 6$ である。以上より、提案手法は一般的な検索手法と同等のスケーラビリティを保つ。

6. 結論・考察

本研究では、画像ベースのローカリゼーション問題をイメージグラフからの回帰問題として定式化を行い、データベースのグラフ構造の利点を用いた、2 ステップの回帰アプローチを提案した。画像ペアを用いた類似度算出のコンセプトは、Street View 画像において良く適合する一方で、triplet や N 対画像への拡張も可能である。最後に、提案手法では画像記述子自体の制約はないため、よりコンパクトな表現である VLAD¹³⁾ 等の利用も可能である。

参 考 文 献

- 1) G.Baatz, K.Koser, G.R., and M.Pollefeys. Handling urban location recognition as a 2D homothetic problem. In *Proc. ECCV*, 2010.
- 2) H.Bay, T.Tuytelaars, and L.VanGool. SURF: Speeded up robust features. In *Proc. ECCV*, 2006.
- 3) C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- 4) O.Chum, A.Mikulik, M.Perdoch, and J.Matas. Total recall II: Query expansion revisited. In *CVPR*, 2011.
- 5) O.Chum, J.Philbin, J.Sivic, M.Isard, and A.Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, 2007.
- 6) M.Cummins and P.Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0. In *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.
- 7) R.Diestel. *Graph Theory*. Springer-Verlag, 2000.
- 8) R.I. Hartley and F.Schaffalitzky. Reconstruction from projections using Grassmann tensors. In *Proc. ECCV*, 2004.
- 9) J.Hays and A.Efros. im2gps: estimating geographic information from a single image. In *CVPR*, 2008.
- 10) <http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>.
- 11) A.Irschara, C.Zach, J.Frahm, and H.Bischof. From structure-from-motion point clouds to fast location recognition. In *CVPR*, 2009.
- 12) H.Jegou, M.Douze, and C.Schmid. On the burstiness of visual elements. In *CVPR*, 2009.
- 13) H.Jegou, M.Douze, C.Schmid, and P.Perez. Aggregating local descriptors into a compact image representation. In *CVPR*, 2010.
- 14) H.Jegou, H.Harzallah, and C.Schmid. A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*, 2007.
- 15) J.Knopp, J.Sivic, and T.Pajdla. Avoiding confusing features in place recognition. In *Proc. ECCV*, 2010.
- 16) F.Li and J.Kosecka. Probabilistic location recognition using reduced feature set. In *Proc. Int. Conf. on Robotics and Automation*, 2006.
- 17) Y.Li, D.Crandall, and D.Huttenlocher. Landmark classification in large-scale image collections. In *ICCV*, 2009.
- 18) Y.Li, N.Snaveley, and D.Huttenlocher. Location recognition using prioritized feature matching. In *Proc. ECCV*, 2010.
- 19) D.Lowe. Local feature view clustering for 3D object recognition. In *CVPR*, pages 682–688. Springer, 2001.
- 20) M.Muja and D.Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*, 2009.
- 21) K.Ni, A.Kannan, A.Criminisi, and J.Winn. Epitomic location recognition. *IEEE PAMI*, 2009.
- 22) D.Nister and H.Stewenius. Scalable recognition with a vocabulary tree. In *CVPR*, 2006.
- 23) J.Philbin, O.Chum, M.Isard, J.Sivic, and A.Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.
- 24) J.Philbin, O.Chum, M.Isard, J.Sivic, and A.Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, 2008.
- 25) J.Philbin, J.Sivic, and A.Zisserman. Geometric latent dirichlet allocation on a matching graph for large-scale image datasets. *IJCV*, 2010.
- 26) A.Roshan and M.Shah. Accurate image localization based on google maps street view. In *Proc. ECCV*, 2010.
- 27) G.Salton and C.Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 1988.
- 28) G.Schindler, M.Brown, and R.Szeliski. City-scale location recognition. In *CVPR*, 2007.
- 29) S.Se, D.G. Lowe, and J.Little. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Intl. J. of Robotics Research*, 21(8):735–758, 2002.
- 30) C.Silpa-Anan and R.Hartley. Localization using an image-map. In *ACRA*, 2004.

- 31) J.Sivic and A.Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.
- 32) N.Snavely, S.Seitz, and R.Szeliski. Photo tourism: exploring photo collections in 3D. In *SIGGRAPH*, 2006.
- 33) P.Turcot and D.Lowe. Better matching with fewer features: The selection of useful features in large database recognition problem. In *WS-LAVD, ICCV*, 2009.
- 34) I.H. Witten, A.Moffat, and T.Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann Publishers, ISBN:1558605703, 1999.
- 35) A.Zamir and M.Shah. Accurate image localization based on google maps street view. In *Proc. ECCV*, 2010.
- 36) W.Zhang and J.Kosecka. Image based localization in urban environments. In *3DPVT*, 2006.
- 37) Y.Zheng, M.Zhao, Y.Song, H.Adam, U.Buddemeier, A.Bissacco, F.Brucher, T.Chua, and H.Neven. Tour the world: building a web-scale landmark recognition engine. In *CVPR*, 2009.