

図 2 各方位に対応するカットアウト画像領域。左下のクエリ画像に対して、同じ水平画角をもつウィンドウをパノラマ画像上でシフトさせていながら、それぞれのウィンドウ領域内から BoF ベクトルを抽出する。各方位に対応する BoF ベクトルが得られるので、位置・方位を同時に推定できる。

- 疎な BoF ベクトルを生成することで、コンパクトかつ計算コストの低い画像表現により、メモリー消費量を節約し処理時間を短縮する。
- 通常の BoF 表現より特徴量の画像における空間的情報を保持するので高精度な検索が可能になる。

1.2 関連研究

画像検索によるアプローチでは、位置情報をもつデータベース画像に対して、撮影された位置情報のないクエリ画像を検索にかけて、その検索結果より撮影された位置・姿勢を推定する。画像検索の精度を上げるための様々な工夫がなされてきた。Chen ら [10] はパノラマ画像のデータベースからカットアウト画像を生成し、さらにそれらを透視投影画像に変換することで、クエリ画像とデータベース画像の間の投影モデルの差異に対応している。しかしカットアウト画像の数に比例してデータベースサイズが増加することになる。また Svetlana ら [5] は従来の BoF モデルでは局所特徴量の画像中における空間的情報を放棄するので、制限された画像表現しかできないと指摘している。Spatial Pyramid Matching と呼ばれる方法で画像を異なるレベルで部分領域に分割し、それぞれの部分領域から特徴量のヒストグラムを抽出して画像を表現している。局所特徴量のグローバル情報ある程度保持できるため、よりリッチな表現となり検索精度を向上できる。本研究でもこのアプローチに倣って、画像を分割表現することで位置・方位推定精度の向上を目指す。

2. BoF モデルによる方位推定実現の課題

位置・姿勢の推定には 6 自由度があるが、本研究では平面の地図上での位置、向いている方位の 3 自由度を前提とする。画像検索にもとづいて位置・方位を推定するもっとも基本的な方法として、図 2 のようにクエリ画像の画角に相当するウィンドウをパノラマ画像上で少しずつずらしながら、それぞれのウィンドウ領域から特徴量と BoF ベクトルを抽出して計算する方法が考えられる。これはパノラマ画像をクエリ画像の水平画角に合わせてカットし、新たにデータベース画像を生成しているに相当する。これら

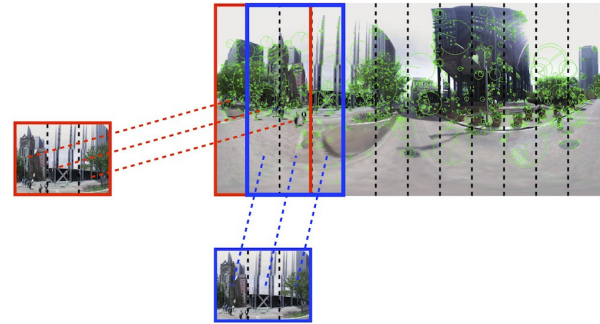


図 3 分割表現によるクエリ画像とパノラマ画像の対応付け。黒の点線のように分割間隔を決めて画像を分割して、それぞれのタイル領域から BoF ベクトルを抽出する。左と下のクエリ画像において、それぞれの対応ウィンドウとのスコアは、対応するタイル領域どうしの内積を独立に計算し、和を取ることで得られる。

のカットアウト画像は独立した画像として見なせるので、データベースサイズが大きくなるが、BoF ベクトルの抽出やスコアの計算方法は従来と変わらない。位置の推定は検索結果における最も類似度の高いパノラマ画像の位置座標を参照すればよいが、方位はカットアウト画像がパノラマ画像のどの範囲に対応するかによって推定される。この方法では位置・方位推定は実現できるが、以下の課題がある。

- ウィンドウのシフト間隔が細かいほど、図 2 のように重なる領域が増え、情報の重複が起きる。
- ウィンドウ全域から一本の BoF ベクトルを抽出するので、非ゼロ要素数が多くメモリー消費量が増える。
- 未知のクエリ画像の画角に対応するためには、あらゆる画角を想定してデータベースを用意する必要がある。これらの問題に対して提案手法では画像の分割表現により解決をはかる。

3. BoF の分割表現を用いた画像検索による自己位置・方位推定手法

3.1 タイル領域にもとづく BoF ベクトル

各方位に対応するカットアウト画像を用いる方法では、膨大なデータベースの用意が必要となる。それに対して提案手法では図 3 のようにあらかじめ決めた分割間隔でパノラマ画像とクエリ画像をタイル領域に分割して、それぞれのタイル領域で BoF ベクトルの抽出を行う。またタイル領域どうしは重ならないので、情報の重複がなくかつ BoF ベクトルがより疎になるため、コンパクトかつ計算コストの低い画像表現となる。さらに分割表現により特徴点の画像上における空間的情報を保持するので、本来の特徴量のマッチングに近づくことができると考えられ、検索精度が向上すると期待できる [5]。

パノラマ画像の水平画角は 360° であり、クエリ画像の水平画角 (Horizontal field of view) を θ 、分割間隔 (interval) を ϕ とすると、パノラマ画像およびクエリ画像から抽出さ

れる BoF ベクトルの本数 m, n は次式のようになる。

$$m = \left\lceil \frac{360}{\phi} \right\rceil \quad (1)$$

$$n = \left\lceil \frac{\theta}{\phi} \right\rceil \quad (2)$$

ここで $\lceil \cdot \rceil$ は天井関数で切り上げを表す。

このようにあらかじめ分割間隔を決めて画像をタイル領域に分割すれば、コンパクトかつ計算コストの低いデータベースを構築することができる。入力されたクエリ画像に対しても同じ分割間隔を適用すればよい。なお本研究ではパノラマ画像を等間隔で分割するが、必ずしもクエリ画像をその間隔で等分できるとは限らないので、 q_n に含まれる領域がわずかだとしてもそのまま計算に用いることとする。

3.2 位置・方位推定のためのスコア計算

提案手法ではあらかじめ分割間隔を決めて画像を分割して BoF ベクトルを用意し、以下のようにクエリ画像と対応するウィンドウ領域とのスコアを計算する。

図 3 のようにクエリ画像の水平面角に対応する幅のウィンドウを考える。クエリ画像を表す BoF ベクトルは $Q = \{q_1, q_2, \dots, q_n\}$ であり、一方 k 枚目のパノラマ画像における i 番目のウィンドウを表す BoF ベクトルは $P = \{p_i, p_{i+1}, \dots, p_{i+n-1}\}$ である。クエリ画像とこのウィンドウ領域のスコアはそれぞれのタイル領域に対応する BoF ベクトルの内積の和であるから、スコア S_{ki} は次のようになる。

$$S_{ki} = q_1 \cdot p_i + q_2 \cdot p_{i+1} + \dots + q_n \cdot p_{i+n-1} \quad (3)$$

i の取る範囲は、 $i = 1 \sim m$ なので、計算する必要のあるスコア $S_{k1} \sim S_{km}$ は、ウィンドウの折り返しに注意すると式 (4) のようになる。

$$\begin{aligned} S_{k1} &= q_1 \cdot p_1 + q_2 \cdot p_2 + \dots + q_n \cdot p_n \\ S_{k2} &= q_1 \cdot p_2 + q_2 \cdot p_3 + \dots + q_n \cdot p_{n+1} \\ &\vdots \\ S_{km} &= q_1 \cdot p_m + q_2 \cdot p_{m+1} + \dots + q_n \cdot p_{m+n-1} \end{aligned} \quad (4)$$

そして $S_k = \max(S_{k1}, S_{k2}, \dots, S_{km})$ より k 枚目のパノラマ画像とのスコアを得る。実際には N 枚のパノラマ画像があるので、 $\max(S_1, S_2, \dots, S_N)$ より検索における最も類似度の高いパノラマ画像を検索する。なお位置推定に関しては本研究では類似度が最大のパノラマ画像の位置情報を割り当てることとする。

一方、方位推定は類似度が最大のパノラマ画像において最も高いスコアを獲得したウィンドウの位置を割り当てる。ここで類似度が最大のパノラマ画像がデータベースにおける k 枚目のパノラマ画像であるとき、最大スコアのウィンドウ位置を $j := \arg\max(S_{k1}, S_{k2}, \dots, S_{kj}, \dots, S_{km})$ とす

ると、推定される方位 ψ は次式のようになる。

$$\psi = (j - 1) \cdot \phi + \frac{\phi \cdot n}{2} \quad (5)$$

このようにして、提案手法では BoF の分割表現を用いて画像検索にもとづいて撮影者の位置・方位を同時に推定する。

4. 実験

4.1 実験で用いるデータベースとテストセット

本研究では評価実験のデータベースとして Pittsburgh 一部の地域で、およそ 10m ~ 15m の間隔で撮影された 2,577 枚の Google Street View パノラマ画像を用いる。検索にかけるクエリ画像のテストセットとしては、Google Street View より密な間隔で撮影した Google Street View Pittsburgh Research Data Set *1 からランダムに選択した 1,000 枚のパノラマ画像を用いる。なおすべての画像には位置・方位情報があり、位置推定実験の評価で用いる。クエリ画像として透視投影画像を Google Street View の投影モデルへ投影した画像 [18] を想定しているが、すでにクエリ用のパノラマ画像があるので、カットアウト画像を生成して投影後のクエリ画像と見なす。なおカットアウトではピッチ角を 15° とし、向いている方位角は $0^\circ \sim 360^\circ$ の範囲でランダムに定める。また Google Street View 画像と Research Dataset 画像の GPS 位置座標のずれを考慮して、以下の実験では推定された位置がクエリ画像の半径 20m 以内なら正解とする。ここで以下の実験で比較する手法をまとめる。

- I. **BoFpanorama** : BoF モデルにもとづいて一枚の画像から一本の BoF ベクトルを抽出する。パノラマ画像を一枚の画像として扱うので、クエリ画像の方位推定は考慮しない。
- II. **BoFcutout** : パノラマ画像において各方位に対応するカットアウト画像を生成しそれぞれの BoF ベクトルを抽出する。クエリ画像の位置・方位推定を実現するベースライン手法。
- III. **tiledBoF** : 画像を分割しそれぞれのタイル領域から BoF ベクトルを抽出する提案手法。

4.2 コンパクトかつ計算コストの低い画像表現

ベースライン手法と比べて、提案手法は情報重複のない疎な BoF データベースを構築するため、コンパクトかつ計算コストが低いことを、実際のデータベースサイズと計算時間をもとに考察する。表 1 に示すように従来 BoF モデルを用いる BoFpanorama はそれら 2 つのどの条件にも影響されず、一定のサイズである。また提案手法 tiledBoF は分割間隔にこそ影響されるが、クエリ画像の画角が変化

*1 Provided and Copyrighted by Google.

表 1 データベースサイズ比較. Interval 5° において BoFpanorama のデータベースサイズを基準とする tiledBoF, BoFcutout のデータベースサイズの比率. なおデータベースサイズでは疎行列のため非ゼロ要素数を考慮する.

	BoFpanorama	tiledBoF	BoFcutout					
Horizontal field of view of query θ	—	—	30°	45°	60°	90°	120°	360°
Database size [ratio]	1	1.07	6.29	9.39	12.47	18.61	24.70	72.00

表 2 計算時間比較. Interval 5° のとき, 3 手法における BoF ベクトルが用意できて内積計算により全データベース画像とのスコアが出力されるまでの時間 [sec].

	30°	45°	60°	90°	120°	360°
BoFpanorama	0.128	0.101	0.144	0.111	0.120	0.117
tiledBoF	0.156	0.200	0.249	0.333	0.382	1.177
BoFcutout	0.704	1.023	1.155	1.577	2.376	6.731

しても決めた分割間隔を適応的に変更するのでデータベースは同じのため, 画角に影響されない. これに対してベースライン手法 BoFcutout では, ウィンドウのシフト間隔やクエリ画像の画角に依存する. ウィンドウのシフト間隔が細かいほど BoF ベクトルの本数が増えて, またクエリ画像の画角が広いほど BoF ベクトルがより密になるためデータベースサイズが大きく膨れ上がる. 提案手法 tiledBoF は一枚のパノラマ画像を分割して表現するので, BoFpanorama のサイズに近くなる.

一方計算時間では, 表 1 の結果より提案手法 tiledBoF はベースライン手法 BoFcutout と比較して非ゼロ要素数が少なく, BoF ベクトルがより疎であるので検索にかかる時間も短いはずである. ここで実際にクエリ画像の BoF ベクトルとデータベースの全パノラマ画像に対する BoF ベクトルが用意できて, 内積計算によって全データベース画像とのスコアが出力されるまでの時間を計測した結果を表 2 に示す. 明らかに提案手法 tiledBoF の方がベースライン手法 BoFcutout より計算時間が短いことがわかる. また BoFcutout ではあらかじめクエリ画像の画角に対応するデータベースを用意する必要があり, もしメモリー節約のためオンラインで合成するならば, より計算時間がかかる.

なお表 1, 2 では interval 5° の場合を示しているが, 図 4, 5 で示す他の interval でも同様の傾向, 提案手法の方がコンパクトかつ計算時間が短いことを確認した.

4.3 分割間隔および画角の検索精度への影響

パノラマ画像を対象とする BoFpanorama を除いて, ベースライン手法 BoFcutout と提案手法 tiledBoF において, ウィンドウのシフト間隔や分割間隔は検索精度にどのような影響を及ぼすのかを評価する (図 4, 5). BoFcutout も tiledBoF もクエリ画像の水平画角が広くなるにつれて, クエリ画像が含む情報が増え, データベースとの対応が取りやすくなり, 検索精度が向上したと考えられる. 一方ウィンドウのシフト間隔・分割間隔に関しては異なる傾向が見られた. BoFcutout ではウィンドウのシフト間隔を

変化させても検索精度の変化はほとんど見られなかった. BoFcutout における一つのウィンドウの幅は変わらないため, ウィンドウのシフト間隔によって一枚のパノラマ画像に対する BoF ベクトルの本数は増減するが, BoF ベクトルに含まれる情報量, すなわち非ゼロ要素数は大きく変化しないためと考えられる.

それに対して tiledBoF では図 5 のように分割間隔を広げていくと精度が上がるが, 逆に広げすぎると精度は下がってしまうことが確認された. このことから適当な分割間隔が存在することが考えられる. 提案手法 tiledBoF では, BoF ベクトルの情報量は分割間隔に大きく左右され, 分割間隔が細かくなると BoF ベクトルは疎になりすぎるため, 正しい対応が取りにくくなったと考えられる. Spatial Pyramid Matching [5] でも過分割では悪影響を与えると述べており, 適当な分割間隔が存在すると考えられる.

図 5 において分割間隔 20°, 30° の場合に注目すると, すべてのクエリ画像の水平画角に対して精度が高いことがわかる. 方位推定を考慮した場合, 分割間隔を 30° に取ると, 方位の分解能が 12 段階程度となる. 分割間隔 20° でも十分な検索精度が得られていることから, 本論文では tiledBoF における分割間隔として 20° を提案する.

4.4 検索精度比較

画像検索では検索結果のランキングの上位 k 番目までの検索結果を利用して, 画像間の幾何学的な対応にもとづく geometric verification [13], 検索結果を新たなクエリとして再検索を行う query expansion [11, 12] により精度向上をはかる方法が取られるため, 本実験では検索結果の上位 100 位まで考慮する. 図 6 よりすべてのクエリ画像の水平画角においても BoFcutout と tiledBoF の検索精度が BoFpanorama を上回っていることがわかる. パノラマ画像におけるすべての特徴点情報をまとめて一本の BoF ベクトルを構成するため, 特徴点の空間的情報が完全に失われ検索精度に影響を与えていると考えられる. 一方 BoFcutout と tiledBoF はクエリ画像に対応するウィンドウを考える

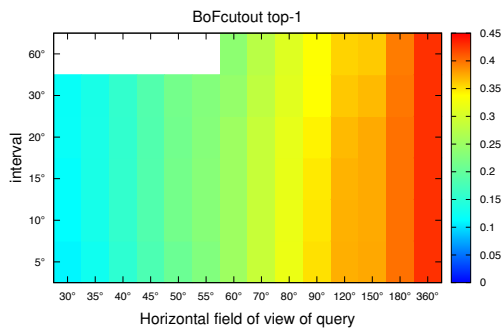


図 4 BoFcutout top-1 の精度比較

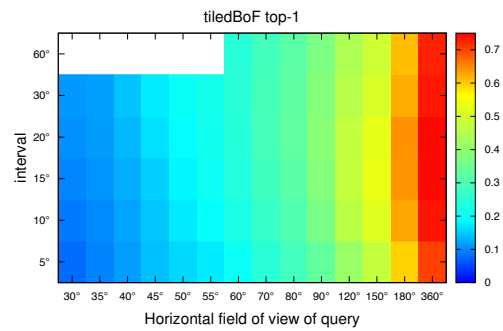


図 5 tiledBoF top-1 の精度比較

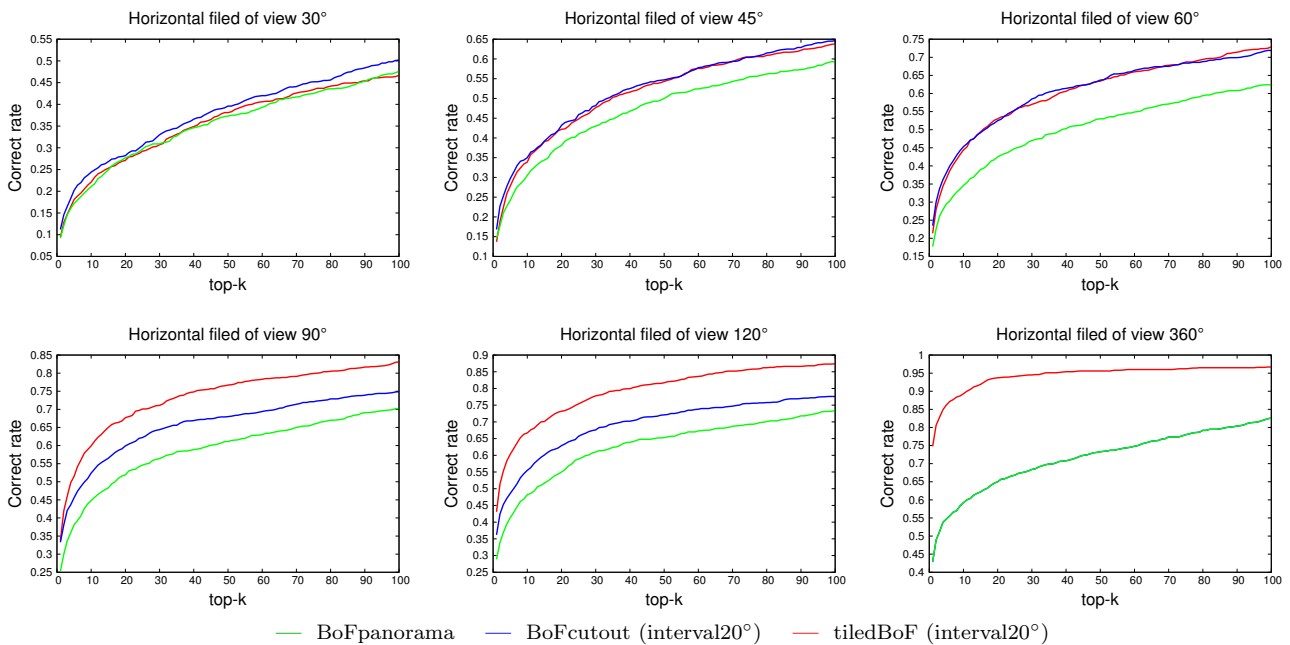


図 6 検索精度の比較

ことにより、ある程度特徴点の空間的情報を残すので、より対応が取りやすく検索精度の向上に貢献することが考えられる。しかし BoFcutout ではまだウィンドウ全域内の特徴点の空間的情報を無視するが、tiledBoF ではさらにそのウィンドウを分割して表現するので、より良い対応が取れることが期待された。画角が小さいときは情報量が乏しく性能を十分に発揮できていないが、60°より大きく画角を広げると BoFcutout を上回る高い検索精度が得られた。またスマートフォンに代表される通常のモバイルデバイスに搭載されているカメラの水平画角がおおよそ 35mm フィルム換算でおおよそ 50° ~ 60° となるが、この範囲において提案手法 tiledBoF はベースライン手法 BoFcutout よりコンパクトな表現であるにも関わらず同等の検索精度を達成できる。

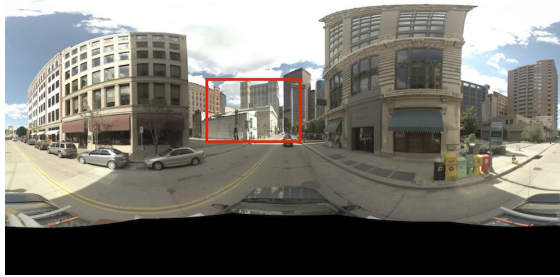
図 7 に実際の提案手法における検索成功例と失敗例を画像で示す。なおクエリ画像の水平画角は 60° であり、分割間隔は 20° とした。また検索ランキング最上位の結果にお

いて、推定される位置がクエリ画像の 20m 以内なら成功とする。失敗例のように壁など特徴的なパターンが少ない場合、また例には示していないが木や空が大部分を占める場合は非常に失敗しやすい。

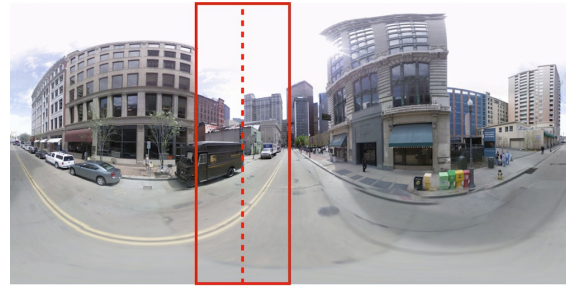
5. 結論

本研究では画像検索のアプローチから位置・方位推定の課題の解決に臨んだ。従来の BoF モデルを応用した方位推定手法 BoFcutout では膨大なデータベースの用意が必要であるのに対して、BoF の分割表現を用いた位置・方位推定手法 tiledBoF を提案した。実験によってベースライン手法 BoFcutout よりコンパクトかつ計算コストが低いことが確認された。また位置推定精度でも同等、画角が広いときはそれ以上の検索精度を得られることを確認した。今後の課題として正しい位置推定ができた場合の方位推定の精度評価に取り組む。

謝辞 本研究は科研費 24700161 の一部助成を受けたも



成功例のクエリ画像



左画像の位置・方位推定結果



失敗例のクエリ画像



左画像の位置・方位推定結果

図 7 検索結果における成功例と失敗例

のである。

参考文献

- [1] Lowe, D.G. : Object recognition from local scale-invariant features. *ICCV*, pp. 1150-1157 (1999).
- [2] Sivic, J.; Zisserman, A. : Video Google: A Text Retrieval Approach to Object Matching in Videos. *ICCV*, pp.1470-1477 (2003).
- [3] Csurka, G.; Bray, C.; Dance, C. and Fan, L. : Visual categorization with bags of keypoints. *ECCV Workshop on Statistical*, pp. 59-74 (2004).
- [4] Nister, D. ; Stewenius, H. : Scalable Recognition with a Vocabulary Tree. *CVPR*, pp. 2161-2168 (2006).
- [5] Lazebnik, S. ; Schmid, C. ; Ponce, J. : Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *CVPR*, pp. 2169-2178 (2006).
- [6] Yunque Li, Noah Snavely, Dan Huttenlocher, Pascal Fua : Worldwide Pose Estimation using 3D Point Clouds. *ECCV*, pp.15-29 (2012).
- [7] Irschara, A. ; Zach, C. ; Frahm, J.-M. ; Bischof, H. : From Structure from Motion Point Clouds to Fast Location Recognition. *CVPR*, pp. 2599-2606 (2009).
- [8] Lowe, D.G. : Distinctive image features from scale-invariant keypoints. *IJCV*, pp. 91-110 (2004).
- [9] Bay,H. ; Tuytelaars,T. ; Gool,L.V. : SURF: Speeded up robust features. *ECCV*, pp.404-417 (2006).
- [10] Chen, D.M. ; Baatz, G. ; Koser, K. ; Tsai, S.S. ; Vedantham, R. ; Pylvanainen, T. ; Roimela, K. ; Xin Chen ; Bach, J. ; Pollefeys, M. ; Girod, B. ; Grzeszczuk, R. : City-Scale Landmark Identification on Mobile Devices. *CVPR*, pp.737-744 (2011).
- [11] Chum, O. ; Mikulik, A. ; Perdoch, M. ; Matas, J. : Total recall II : Query expansion revisited. *CVPR*, pp. 889-896 (2011).
- [12] Arandjelovic, R. ; Zisserman, A. : Three things everyone should know to improve object retrieval. *CVPR*, pp. 2911-2918 (2012).
- [13] Tsai, S.S. ; Chen, D. ; Takacs, G. ; Chandrasekhar, V. ; Vedantham, R. ; Grzeszczuk, R. ; Girod, B. : Fast geometric re-ranking for image-based retrieval. *ICIP*, pp. 1029-1032 (2010).
- [14] Lowe, D.G. : Local feature view clustering for 3D object recognition. *CVPR*, pp. I-682 - I-688 (2001).
- [15] Rabinovich, A. ; Vedaldi, A. ; Galleguillos, C. ; Wiewiora, E. ; Belongie, S. : Objects in context, *ICCV Oral*, pp.1-8 (2007).
- [16] Varma, M. ; Ray, D. : Learning the discriminative power-invariance trade-off. *ICCV Poster*, pp.1-8 (2007).
- [17] Jegou, H. ; Douze, M. ; Schmid, C. ; Perez, P. : Aggregating local descriptors into a compact image representation. *CVPR*, pp. 3304-3311 (2010).
- [18] Petr Gronát, Michal Havlena, Josef Šivic, Tomáš Pajdla, : Building Streetview Datasets for Place Recognition and City Reconstruction. <http://baikal.graphics.cs.cmu.edu/cdoersch/im2gps2/petr/Report/Gronat-TR-2011-01.ps>