Efficient Panorama Database Indexing for Indoor Localization

Jean-Baptiste Boin
Stanford University
jbboin@stanford.edu

Dmytro Bobkov
Technical University of Munich
dmytro.bobkov@tum.de

Eckehard Steinbach
Technical University of Munich
eckehard.steinback@tum.de

Bernd Girod
Stanford University
bgirod@stanford.edu
Indoor Localization

- Task: panorama retrieval using a single query image

- Goal is fast coarse localization; can be used as a first pass for a more complex fine localization system

- Query/database asymmetry
System design

View sampling → Descriptor extraction → Descriptor pre-processing → Indexing
Retrieval with descriptor aggregation

WHAT DESCRIPTORS to aggregate?

WHAT AGGREGATION METHOD to use?

HOW TO USE aggregated descriptors?
Contributions

- Systematic evaluation of view sampling and aggregation
  - Fine sampling of panoramas + descriptor aggregation is preferred to coarse sampling
  - Pooling descriptors with Generalized Max Pooling (GMP) is superior to mean pooling
- Speed up search with hierarchical index based on the location and orientation of the views
Background – Image representation

- Traditional pipeline: hand-crafted features
  - Local patch representation: SIFT [Lowe, ’04]
  - Global descriptor:
    - Bag of Words (BoW) [Sivic et al., ’03]
    - Fisher Vectors [Perronnin et al., ’07]

- CNN-based features
  - Representations extracted from networks trained on other tasks
  - Can be fine-tuned for improved results
Background – Descriptor aggregation

- Generalized max-pooling (GMP) [Murray et al., ’14]
  - Increased similarity to ALL descriptors
Background – Indexing and search

- Task: Nearest Neighbor (NN) search
  - Database: $X = \{x_1, \ldots, x_N\}$, with $x_i \in \mathbb{R}^d$, $\|x_i\|^2 = 1$
  - Query: $q \in \mathbb{R}^d$, $\|q\|^2 = 1$
  - Find $i$ that maximizes $q^T x_i$
  - Exhaustive search: $O(Nd)$

- High dimensional exact NN search is hard
  - When $d \geq 10$, no gains compared to exhaustive search [Weber et al., ’98]

- Approximate Nearest Neighbor (ANN) techniques:
  - Space-partitioning techniques: FLANN [Muja et al., ’14]
System design

View sampling
Descriptor extraction

Descriptor pre-processing

Indexing
View sampling

- Rationale: higher similarity when matching with limited FoV queries
- Vertical sampling
  - Sampled at elevations -30°, 0°, 30°
- Horizontal sampling
  - Sampling rate of 48 (step = 7.5°)
  - Considerable overlap between views
- 144 views per panorama
System design

View sampling ➔ Descriptor extraction ➔ Descriptor pre-processing ➔ Indexing
Descriptor extraction

Sampled view

Deep Image Retrieval
[Gordo et al., '16]

Centering + L2-normalization

Image descriptor
Evaluation

- Similarity between query and database descriptors computed from L2 distance (order is equivalent to cosine similarity)

- Convert list of views to list of panoramas by keeping the first occurrence of each panorama

- Evaluate average precision for the query
Datasets

- **WUSTL Indoor RGBD dataset**
  [Wijmans et al., ’17]
  > 129 geo-localized panoramas captured in a university building

- **Matterport3D dataset**
  [Chang et al., ’17]
  > Used as distractors
Datasets

- **InLoc** dataset [Taira et al., ’18]
  - 329 queries
  - Same building as the WUSTL dataset
  - Captured at a different time and on a different camera (iPhone)
  - 6DoF manually verified reference poses

- Total of 693 panoramas per query
Results (baseline)

Faster search

Higher accuracy

Baseline (all sampled views, exhaustive search)
System design
Descriptor pre-processing

- Rationale: reduce effective size of database (number of descriptors compared per query) while keeping the performance high

(4) Sub-sampling
(4) Aggregation
Descriptor pre-processing

Elevation

Azimuth

Aggregation (1,1)
Descriptor pre-processing

Aggregation (8,1)

elevation

-30°
0°
30°

azimuth

0°
90°
180°
270°
Descriptor pre-processing

Aggregation (4,3)
Descriptor pre-processing

- **elevation**
  - 30°
  - 0°
  - -30°

- **azimuth**
  - 0°
  - 90°
  - 180°
  - 270°

Sub-sampling (8,1)
Descriptor pre-processing

Sub-sampling (4,3)
Results (pre-processing)

- Aggregation offers better trade-offs than sub-sampling
- GMP is preferable to MP when aggregating many dissimilar descriptors
System design

View sampling  Descriptor extraction  Descriptor pre-processing  Indexing
Indexing

- Hierarchical aggregation: best of both worlds
  - Upper levels: coarse search = large complexity gains
  - Lower levels: fine search = higher retrieval performance
- Node: set of database descriptors; leaf: single database descriptors
- Index search:
  - Compute distance of a query with all children of the root
  - Pick node with lowest distance; put other nodes in priority queue
  - Continue until reaching leaf node
  - Pull node with lowest distance in priority queue and recurse
- Early stopping: allows exploration of cost/accuracy trade-off
Indexing

- Data-based hierarchy (DBH)
  - Based on k-means tree algorithm in FLANN [Muja et al., '14]
    - Choose branching factor B
    - Recursive k-means until each cluster contains < B descriptors
  - Internal node descriptors:
    - Pooled with GMP
    - Normalized

- Geometry-based hierarchy (GBH)
  - Based on view orientation
  - + room-level aggregation
Results (DBH)

- Modifications to FLANN are critical for an acceptable performance.
Results (GBH)

- Importance of pre-processing (last stage of the hierarchy)

Technical University of Munich
Results (Summary)

- 374 descriptors/query
- ~50k descriptors/query
- 133x complexity reduction
- 44x
- 3x

mAP (%) vs. #descriptor comparisons

- Sub-sampling (exh. search)
- Aggregation (exh. search)
- DBH
- GBH

Technical University of Munich
Conclusions

- Reducing database size through pre-processing by aggregating neighboring views (3x speed increase)
- GMP provides a better representation for a set of descriptors
- Faster search by nesting multiple aggregation levels (44x speed increase)

Code available on GitHub:
https://github.com/jbboin/panorama-indexing-localization
Questions?